NET WORK LISER

Discussion Paper Series

IZA DP No. 18584

April 2026

Workforce Quality and Early Childhood Development at Scale

Sarah Cattan
IFS, HCEO, IZA@LISER and CESifo

Gabriella Conti
University College London, IFS, CEPR, CESifo, CSEF, IZA@LISER and RFBerlin

Christine Farquharson
Institute for Fiscal Studies

IZA DP No. 18584 April 2026

Workforce Quality and Early Childhood Development at Scale*

Abstract

Early childhood programmes frequently lose effectiveness at scale, yet the role of the workforce remains poorly understood. We document substantial heterogeneity in workforce effectiveness in England's national home-visiting programme for first-time teenage mothers, despite a highly-structured curriculum and well-qualified staff. Exploiting quasi-random assignment of mothers to family nurses, we estimate that a one standard deviation increase in workforce effectiveness raises children's cognitive and socio-emotional development by 0.20-0.23 SD. Structural quality — observable worker characteristics — does not predict effectiveness, but process quality — how visits are delivered — does. Greater effectiveness is linked with improvements in maternal mental health and risk behaviours.

JEL classification

I18, I38, J13, J24

Keywords

early childhood development, home visiting, workforce quality, process quality, scaling, Family Nurse Partnership

Corresponding author

Gabriella Conti
gabriella.conti@ucl.ac.uk

* *Acknowledgements*: The authors gratefully acknowledge funding from the Tavistock and Portman NHS Foundation Trust, the ESRC Centre for the Microeconomic analysis of Public Policies at the Institute for Fiscal Studies, and the NORFACE DIAL for the project "Growing up Unequal? The Origins, Dynamics and Lifecycle Consequences of Childhood Inequalities". Conti acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No.819752 DEVORHBIOSHIP – ERC-2018-COG), and from The Leverhulme Trust, via a Philip Leverhulme Prize. We are indebted to Ruth Rothman, Phil Howlin and Andreea Moise from the Family Nurse Partnership National Unit for answering our numerous questions about the data and the programme. We thank seminar participants at the "Medical Interventions and Socioeconomic Outcomes" Workshop at Odense University, UCL Research Day Workshop, EALE, EEA, ESPE conferences, the Economic Demography Workshop, the Paris School of Economics workshop on early childhood development, the Workshop on Labour and Family Economics at the University of York, the NORFACE-DIAL conference, the Barcelona Summer Forum, IBEO Alghero, NBER Summer Institute Children Group, Bank of Italy Human Capital Workshop, Workforce Workshop at the University of Surrey, Opportunity & Inclusive Growth Institute, and at the University of Copenhagen, Venice Ca' Foscari, Queen Mary, Bologna, Padova, Princeton, Naples, Oxford, Warwick, Aarhus, LSE, Zurich, Bocconi, PUC, University of Chile, UCLA, UCSB, NYUAD, York, Bergen, Berlin, CESR USC, UC Davis, Rotterdam, Hitotsubashi, Copenhagen Business School, CESifo and the Institute for Fiscal Studies for useful comments. All errors remain our own.

1 Introduction

Inequalities in health, cognition, and socio-emotional development emerge early in life and tend to persist without intervention. A large literature shows that early childhood investments can yield high social returns, making them key to anti-poverty and social mobility policy (Almond et al., 2018; Hendren and Sprung-Keyser, 2020). Yet a growing body of evidence shows that interventions which produce positive results under trial conditions often fail to replicate those impacts when implemented at scale (List et al., 2021; Duncan et al., 2023; List, 2026). Understanding why programme effectiveness attenuates at scale is now a first-order challenge for the design of effective social policy.

To date, research on what governs the effectiveness of early years programmes at scale has largely focused on the demand side: which children or families benefit most, and how targeting can be improved. There is also some evidence on demand-side mechanisms, such as how programmes change parental investments (Attanasio et al., 2015). Considerably less attention has been paid, however, to *supply-side* determinants, and in particular to the role of frontline delivery. This gap is striking, because staff time constitutes the primary cost in most early years programmes (and indeed in many public services). And the practitioner-client relationship is the core technology through which programme content is translated into behaviour change, making workforce effectiveness a first-order consideration for effective interventions.

A flourishing literature in other domains of public service delivery shows that the quality of service providers and frontline workers can be a powerful determinant of outcomes. In education, teacher value-added measures have been shown to predict both test score and non-test score outcomes, including adult earnings (Chetty et al., 2014a,b; Jackson, 2018; Petek and Pope, 2023), while classroom observations reveal that it is the quality of teacher–student interactions (rather than observable credentials) that drives learning gains (Araujo et al., 2016). In healthcare, recent work has quantified substantial variation in physician effectiveness, with task-specific skills that interact with patient-specific risks (Zaranko et al., 2023; Currie and Zhang, 2024; Dahlstrand, 2025; Stoye, 2025). In child protection, the identity of the assigned caseworker can be decisive for children's outcomes (Doyle Jr, 2007; Bald et al., 2022). Across these settings, the literature consistently finds that individual providers make a substantial difference. However, it remains limited in characterising what makes one service provider more effective than another. In the few instances where this question has been studied, a consistent finding is emerging: observable credentials and qualifications explain little of the variation in effectiveness, whereas measures of what providers actually do (the quality of interactions, adherence to best practice, and task-specific clinical skills) are far more predictive. Whether this pattern extends to early childhood programmes is an open and important question. In such programmes, delivery unfolds over months or years; the worker's effect on child outcomes is largely indirect (operating through changes in parental behaviour); and effectiveness is inherently multi-dimensional.

In early childhood programmes specifically, an emerging literature has begun to study the role of quality, drawing on a long tradition in education research distinguishing between structural quality (e.g. staff-to-child ratios, staff qualifications) and process quality (e.g. teaching behaviours, the quality of adult–child interactions). In centre-based settings, Blau and Currie (2006) provide an early overview; Blanden et al. (2017) show that staff qualifications have a weak association with child outcomes within nurseries; Andrew et al. (2024) find that a low-cost training intervention improved teaching quality and children's cognitive development in Colombian preschools; and Ganimian et al. (2024) show that increasing preschool staffing in India improved test scores via increased instructional time. In home visiting, a distinct and common type of early childhood programme, the evidence base is thinner: Leer and Lopez-Boo (2019) review programme design features, and a few recent papers study visit quality and delivery in small-scale trials (Conti et al., 2024, 2025; Schepan et al., 2025; Wüst and Sørensen, 2025). To date, however, no study has provided causal evidence on the contribution of the frontline worker within a *scaled* early childhood programme - whether centre-based or home-visiting.

This paper fills this gap. We study the role of workforce quality in driving the effectiveness of England's Family Nurse Partnership (FNP), a major large-scale national home visiting programme for first-time teenage mothers. Adapted from the well-studied U.S. Nurse Family Partnership, the FNP offers up to 64 structured home visits delivered by a specially trained family nurse (FN), from early pregnancy through the child's second birthday. The programme has been experimentally evaluated, with lasting positive effects on child development (Robling et al., 2015, 2022). It operates at scale, reaching around 10,000 mothers (over half of its target population) at its peak in 2016; and it is delivered by degree-educated, experienced health professionals who are also given programme-specific training and follow a structured curriculum. These institutional features make the FNP a particularly stringent test for whether *who* delivers still matters when *what* is delivered is tightly specified.

Two unique strengths of our research design help us overcome the challenges that have prevented previous work from credibly identifying workforce effectiveness in early childhood programmes. First, we exploit the quasi-random assignment of mothers joining the FNP to a family nurse: new clients are allocated to family nurses by supervisors primarily to equalise caseload size and complexity within their teams at the time of referral; crucially, this allocation occurs before the client is onboarded and undergoes a detailed baseline interview, and is based only on a small set of referral-time variables that are fully observable to us: age, gestational age, language, and ethnicity. We validate this institutional programme feature empirically by showing that - conditional on this small set of maternal characteristics, measures of family nurse caseload size and complexity, month of enrolment, and programme site - there is no systematic association between a large set of maternal characteristics (collected during the baseline interview) and a variety of family nurse characteristics. This evidence, consistent with the quasi-random allocation of mothers to family nurses, enables us to identify the distribution of individual family nurse effects on the outcomes of

children they visit.

The second strength of our research design is the exceptionally rich longitudinal information collected about the characteristics and outcomes of mothers, children, family nurses, and supervisors - as well as the interactions between them. Our sample covers about 32,000 mother–child dyads, 1,300 family nurses, one million visits, and approximately 2,700 accompanied (supervisory) visits over the programme's first decade of operation. This not only enables us to estimate and compare family nurse effectiveness on a range of outcomes, but also to explore how it correlates with close to 40 different family nurse characteristics and behaviours, and what maternal behaviours relatively 'good' family nurses activate.

We start by estimating family nurse effectiveness at improving five child outcomes: health at birth; and cognitive and socio-emotional development at 12 and 24 months, measured using the validated Ages and Stages Questionnaires (ASQ-3 and ASQ-SE). Our preferred specification identifies family nurses' effectiveness within sites by regressing site-demeaned outcomes on FN fixed effects, basic maternal characteristics and measures of caseload size and complexity - all recorded at the time of assignment - and a month-of-enrolment fixed effect.[1] Because family nurses have numerically uneven caseload sizes, we apply an empirical Bayes shrinkage estimator to reduce sampling noise for those with small caseloads.

This approach, which exploits the quasi-random allocation of clients to family nurses within local FNP teams, identifies the distribution of workforce effectiveness in a model where both family nurse and site-specific factors affect children's outcomes additively. That is, it strips out measures of family nurse effectiveness from site-specific factors that may impact children's outcomes, such as the quality of supervision, team composition and local services. We complement this with estimates of the distribution of family nurse effectiveness based on non-demeaned outcomes. Since family nurses rarely switch sites, we cannot separately identify family nurse and site fixed effects. This approach therefore recovers the distribution of effectiveness across sites, where each estimated effect additively captures both a family nurse-specific and a site-specific component. Combined with our estimates of the within-site variation in family nurse effectiveness (which difference out the common site component), this allows us to bound the contribution of site-specific factors to the total across-site dispersion.

We document three sets of results. First, we show that there is substantial heterogeneity in family nurse effectiveness within sites. A one standard deviation (SD) increase in family nurse quality raises cognitive development by 0.20 SDs and socio-emotional development by 0.23 SDs at age two. These effects are 1.7 to 2.8 times larger than what would be expected if family nurses had no systematic effect on outcomes, as assessed by randomly permuting family nurse assignments

[1] New clients are allocated to family nurses by the team supervisor, primarily to equalise caseload size and complexity across team members. We conduct our demeaning and our quasi-random assignment test at the site level, which is the level at which governance and allocation rules are set, and which is equivalent to the team level in more than 80% of the sites.

within sites. Put differently, the difference between being assigned a more versus a less effective family nurse is of the same order as the programme's own experimental impact on cognition at age two (Robling et al., 2015). That we find such dispersion under a structured model with trained, relatively well-compensated staff indicates that the individual worker's characteristics, traits, and behaviours still play a very central role, even under tight protocols.

The rest of the paper builds on this core result to better understand what distinguishes better from worse family nurses. First, we find that *comparative advantage* across developmental domains, rather than absolute advantage, best describes this workforce. While effectiveness at improving different dimensions of child development is positively correlated, these correlations are not especially large. For example, a family nurse who is one standard deviation more effective at promoting cognitive skills in a 1-year-old is only 0.38 SD more effective at promoting socio-emotional skills at age 1 (but she is almost 0.7 SD more effective at promoting cognitive skills at age 2).

Second, exploiting the rich information that we have about family characteristics and practice styles, we find that *how* family nurses conduct visits is far more predictive of their effectiveness than their observable characteristics. Even with detailed measures of demographics, qualifications, and training, structural quality indicators predict little of the variation in family nurse effectiveness across child outcomes. By contrast, process indicators — what happens during visits — matter more. We find that longer visits, higher coverage of planned content, and higher maternal engagement and understanding are associated with greater improvements in child development. The allocation of time across the curriculum also matters, with relatively more attention to the home environment and sustained engagement predicting larger developmental gains. Supervisor 'implementation' ratings add little additional explanatory power, partly due to ceiling effects.

Finally, the mechanisms through which effective family nurses operate align with the programme's theory of change and differ by developmental domain. Constructing factor indices along five channels — maternal risky behaviours, socio-economic circumstances, relationships, mental health, and investments in the child — and using leave-out measures of family nurse effectiveness constructed following standard methods in the value-added literature (Chetty et al., 2014a) to avoid mechanical correlation, we find distinct pathways: family nurses who are particularly effective at promoting cognitive development operate primarily by reducing maternal risky behaviours during pregnancy and the early years, while those who are more effective at promoting socio-emotional development operate primarily through improvements in maternal mental health. To our knowledge, this is the first provider-level evidence on how workforce quality translates into child development in a scaled early years programme.

Our paper makes key contributions to two literatures. First, we advance understanding of what determines the effectiveness of early childhood programmes at scale. Previous work has asked whether programme quality matters, but as the Duncan et al. (2023) handbook chapter concludes, existing research "falls short of answering fundamental questions about what works, for whom, and why." Our paper provides the first worker-level evidence on the "why", from the supply side.

We show that, even in a highly structured programme with well-qualified staff, the individual worker is a crucial determinant of programme effectiveness, and that it is implementation quality, not observable credentials, that distinguishes more effective workers from less effective ones. Our finding that family nurse effectiveness is domain-specific further implies that workforce policies in early childhood (hiring, training, allocation, and supervision) should aim not only to raise average quality, but also to match workers' comparative strengths to families' specific needs, just as recent work in healthcare has argued for provider–patient matching on the basis of task-specific skills (Dahlstrand, 2025).

Second, we contribute to the broader literature on workforce quality in public service delivery. Value-added approaches have been extensively developed in education (Koedel et al., 2015; Chetty et al., 2014b) and are increasingly applied in healthcare (Currie and Zhang, 2024; Dahlstrand, 2025; Stoye, 2025), but their extension to other settings remains limited. Our paper applies a similar worker fixed effects approach to a context where the worker's influence on children is largely indirect - through sustained changes in maternal behaviour over repeated interactions - and where the outcome is multi-dimensional. By combining a credible quasi-experimental design with very rich data on both worker characteristics and delivery processes, we are able not only to quantify the importance of the workforce but also to characterise what effective workers do differently and through which channels they operate. These findings suggest that greater attention to workforce quality and delivery processes, informed by the kind of multi-dimensional provider analysis that is now gaining traction in healthcare and education, may be key to successfully scaling early childhood interventions.

2 The FNP programme

The Family Nurse Partnership (FNP) is a home visiting programme aimed at improving the outcomes of first-time teenage mothers and their babies. It is the UK adaptation of the U.S. Nurse Family Partnership, which has been shown (via randomised controlled trials in three different sites) to have consistent and enduring benefits for maternal and child health.[2]

Its objectives are to improve the outcomes of pregnancy by helping women to promote their prenatal health; to improve children's health and development by enabling mothers to provide more

[2]The Nurse-Family Partnership (NFP) has been tested in three flagship US randomised trials showing improvements in maternal life-course outcomes and child health (see Olds et al. (1986) and Olds et al. (1986) for short-term impacts in the Elmira trial; Kitzman et al. (1997, 2000); Olds et al. (2004, 2007); Kitzman et al. (2010) for short-term and 3,6,9, and 12-year follow-ups in the Memphis trial; and Olds et al. (2002, 2014) for short-term and 3 and 6-year follow-ups in the Denver trial). Long-run follow-ups report sustained benefits into adolescence/early adulthood, including reductions in youth criminal/antisocial behaviour at age 15 (Olds et al., 1997, 1998) and improved life-course outcomes at age 19 in the Elmira trial (Eckenrode et al., 2010) and at age 18 in the Memphis trial (Olds et al., 2019; Kitzman et al., 2019; Conti et al., 2024). More recent randomised evidence from the largest trial of the NFP to date, conducted at scale in South Carolina, finds no significant effects on adverse birth outcomes (McConnell et al., 2022), child health outcomes (Swanson et al., 2025), or birth spacing (Steenland et al., 2025). The trial's authors note that the use of administrative data preclude observation of child development or health status, and call for future work on longer-term developmental outcomes — precisely the outcomes for which the Building Blocks trial found the strongest and most lasting effects (Robling et al., 2015, 2022)

competent care for their children; and to improve women's life course by supporting them to plan any subsequent pregnancies, finish their education and/or find employment. The programme offers first-time teenage mothers up to 64 regular home visits with family nurses covering a structured curriculum, starting during pregnancy and continuing onto the child's second birthday. These Family Nurses (FN) are highly qualified: they are required to have nursing or midwifery qualifications; to be registered with the Nursing and Midwifery Council; and to be educated to degree level.

The FNP in England was experimentally evaluated in the 'Building Blocks' trial, where eligible mothers were randomly assigned to either the FNP or standard health visiting care universally available through the National Health Service (NHS). No significant impacts on the trial's pre-designated primary outcomes (birth weight, smoking, repeated pregnancy and hospital admissions) were found (Robling et al., 2015).[3] However, the trial did find significant and lasting benefits for child cognitive and language development. Specifically, children from families visited by a Family Nurse were significantly less likely to have cognitive development concerns reported by their mothers at age 2; they also showed higher language skills at 12, 18 and 24 months. These cognitive impacts persisted at primary school entry, when treated children were 6 percentage points more likely to achieve a good level of development at age 5, and 5 percentage points more likely to reach the expected academic standard at age 7 (Robling et al., 2015, 2021). While the results of the RCT motivate our focus on child outcomes, our objective in this paper is not to evaluate the impact of the programme relative to standard care. Rather, we aim to understand the role that workforce effectiveness plays in promoting child outcomes within a programme implemented at scale.

Allocation of clients to family nurses in the FNP programme

In the FNP the administrative unit of organisation is the site, which corresponds to a FNP licensed operational catchment/commissioning unit, often (but not always) mapping one-to-one to a Local Authority area (Barnes et al., 2008). Within each site, service delivery is organised into (one or more) local FNP teams, each led by a clinical supervisor. In most sites in our sample period, there is a single operating supervisory structure at any point in time; where we observe overlapping supervisors, this reflects short handovers or temporary cover, job-share arrangements, and/or parallel teams in larger sites (consistent with the licensed model's cap on the number of family nurses per supervisor).[4] Regardless of internal supervisory structure, governance and allocation rules are at the site level, making the site the natural unit at which to condition to align our identification with

[3] In 2010, the programme was also extended to cover Scotland and Northern Ireland; however our data only cover its large-scale implementation in England.

[4] Using supervisor tenure start and end months, we construct a site-by-month panel and count the number of supervisors simultaneously active in each site-month. In 67.4% of sites, there is never more than one supervisor active in the same month. A further 14.0% of sites exhibit exactly one month with two or more supervisors active, consistent with short handovers or temporary cover. The remaining 18.6% of sites have two or more months with overlapping supervisors, consistent with job-share arrangements and/or the operation of multiple teams within larger sites (noting that, under the licensed model, a supervisor's team is capped at roughly eight family nurses). For sites with sustained overlap, we cannot determine from the data whether this reflects a job-share arrangement or a parallel team structure, as the FTE data are incomplete; external documentary evidence suggests that both arrangements are common.

the institutional assignment mechanism.[5]

Potential clients interested in joining the FNP pass through four stages before they eventually start working with a family nurse. This sequence of events, and what is known/not known by the programme about the mother at each stage, play a key role in underpinning our empirical design, so we describe it in detail here. In section 4, we also provide empirical evidence to suggest that the actual allocation of mothers to family nurses is in line with these principles.

The first stage is the **referral** stage: mothers can self-refer, but in most instances they are referred to their local FNP team by their midwife or doctor. In all cases, they fill out a referral form containing basic information: contact details, hospital of reference, date of birth, expected due date, language, ethnicity, and residential status. The expectant woman is then assessed against the strict FNP eligibility criteria: mothers must have been aged 19 or under at last menstrual period; have a first pregnancy confirmed by health services (those whose previous pregnancies ended in miscarriage, termination or stillbirth are also eligible); be less than 28 weeks and 6 days of gestation; live within the catchment area covered by the local FNP team; and have no plans to place the child for adoption at enrolment. Non-English speaking clients and clients with learning disabilities are also eligible.

The second stage is the **invitation** stage: if the potential client meets the eligibility criteria, she is then invited to join the FNP. The third stage is the acceptance and **allocation** stage: if the potential client accepts, then the FNP supervisor allocates her to a family nurse in her team.[6] FNP National Unit (2017) stipulate that this allocation decision should be driven by family nurses' availability: more specifically, new clients are allocated to equalise the 'complexity-adjusted' caseloads of family nurses within the team.[7] This caseload-driven allocation is also adjusted for a couple of specific factors (collected on the referral form): the client's ethnicity and preferred language. These are the only two characteristics for which matching is purposefully sought by the supervisor. Importantly, aside from these characteristics included in the referral form, the supervisor has no other information about the prospective client at the time of allocation.

The fourth and last stage is the **enrolment** stage: when the allocation is finalised, the client and her family nurse together complete the enrolment paperwork. At this point, the FN collects extensive additional information about her new client, including socio-economic demographics, health status and risk factors (such as smoking, drinking or experience of domestic violence). After this, the

[5]During the period covered by our data (enrolment between 2007–2013), the FNP was commissioned through Primary Care Trusts (PCTs), which were abolished on 1 April 2013 under the Health and Social Care Act 2012. Programme regulations, eligibility criteria, and the curriculum were set centrally by the FNP National Unit, meaning that site-level variation in our data reflects differences in implementation and workforce rather than differences in programme rules.

[6]In subsection 4.3 we conduct our quasi-random assignment test at the site level, which is the level at which governance and allocation rules are set, and which is equivalent to the team level in the 81.4% of sites that never had more than one month of simultaneous supervisor overlap.

[7]This is consistent with the evidence in Figure A.1: at the beginning of their tenure in a site, family nurses have a lower stock of clients and receive higher inflows of new clients, until these stabilise.

home visits start.

This four-stage process implies that, while the FNP programme collects detailed information about clients from enrolment onwards, the supervisor knows very little about the mother at the time of assigning her to a family nurse. As we argue in section 4, this feature underpins our argument that mothers are quasi-randomly allocated to family nurses.

3 Data and descriptive statistics on FNP implementation

3.1 The FNP-IS data

From its inception, the FNP programme collected very rich information on its clients, workforce and implementation through the purpose-built FNP Information System (FNP IS). All the analysis in this paper is based on these data. While we have data available for the first decade of the programme (2007 to 2017), we restrict our analysis to clients and family nurses who joined the FNP before 2014: this allows us to observe children through to their second birthday (i.e., at the end of the programme). This sample restriction also corresponds to the period of most faithful programme implementation: following the publication of the Building Blocks trial evaluation results in 2015 (Robling et al., 2015), some FNP sites introduced adaptations to the programme model.

The FNP IS records extremely detailed information on clients, family nurses, and visit delivery. On the client side, the FNP IS dataset contains detailed information on mothers' characteristics, behaviours, and outcomes at various points through the programme. These include mother's physical and mental health (including health conditions, height and weight); health habits (smoking, drinking and drug use); relationship status (including history of domestic violence); and demographics (including household composition, education, employment and income, and use of health and social care services). The dataset also contains rich information on children's health at birth and their cognitive and socio-emotional development at four different points in the programme. We will provide more information about the specific measures we use as key outcomes in subsection 4.6.

On the delivery side, the FNP IS data contains detailed information about the characteristics of the workforce: ethnicity, age, qualifications, training and experience of the family nurse prior to joining the FNP programme. Following the early childhood literature, we refer to these observable characteristics as structural quality indicators, since they are the kind of information typically available to an employer at the point of hiring.

We also have two sources of process quality data — information on how visits are actually delivered, which is harder to observe at the point of hiring but potentially more informative about effectiveness. The first source is *monitoring data* recorded by the FN herself: after every home visit she conducts, she records the visit duration, the share of the planned content which was covered, and the proportion of time spent on each of the five programme content strands. Additionally, the FN rates on three five-point scales the mother's engagement during the visit, and her understanding

of, and conflict with, the material. Importantly, the family nurse is not asked to rate her own performance directly, but is instead prompted to record a mix of objective measures and assessments of how the visit went *for the mother*.

The second source of process quality data is taken from the *supervisors' evaluations* of the FN. As discussed in section 2, supervisors accompany the FN on one or more of her home visits and assess her against the Visit Implementation Scale.[8] This assessment thoroughly evaluates all aspects of preparing for and conducting the visit, including: planning/scheduling; greeting and relating to the client; reviewing previous content and goals; assessing the client's current status; delivery of planned curriculum; and goal setting for the next visit.

3.2 Implementation of the FNP in England

The FNP-IS allows us to create a precise picture of the programme's implementation in England. Relative to the wide range of early years programmes in operation across Local Authorities (LAs) in England, the FNP stands out for its targeting; its highly qualified workforce; its intensive and structured curriculum; and the scale of its operation. We describe each of these aspects in turn.

Table 1 describes the client base of the FNP during our period of analysis and shows that the eligibility criteria were for the most part respected. In line with the programme guidelines, mothers who enrolled between 2007 and 2013 were, on average, 17.8 years old at enrolment; fewer than 5% were older than 19.[9] On average, mothers enrolled at 18 weeks' gestation; 95% enrolled before 28 weeks and 6 days, consistently with the programme's eligibility threshold. Table 1 also documents the high levels of social and economic disadvantage among FNP clients: for example, over half of mothers (and a third of those aged 16 or younger) are not in employment, education, or training; half of mothers are receiving at least one type of service (e.g., support for substance abuse or for domestic violence); and 14% are registered in the social care system.

Family nurses carry a maximum caseload of 25 clients, but - as Figure A.4 shows - such high caseloads are unusual: the median family nurse has 14 clients at any given time and carries out 24 visits per month (just under two per client). They work in a specific FNP site, as part of a team of about 8 family nurses led by a supervisor, who is typically a more experienced FN (see Figure A.5).[10] In addition to carrying a small caseload of their own clients, supervisors are responsible for the management and support of the FNs in their team. This includes attending home visits with each FN to provide an assessment of her performance and to give her detailed feedback.[11]

[8] These supervisory visits were intended to occur roughly every six months, but in practice we find considerable heterogeneity in their frequency and timing. However, we find no evidence that such timing is predicted by either the FN characteristics or those of her clients.

[9] A small number of LAs made limited exceptions to the eligibility age where there were unfilled places on the programme.

[10] As previously reported, the site level is equivalent to the team level in the 81.4% of sites that never had more than one month of simultaneous supervisor overlap.

[11] Correspondence with the FNP National Unit has clarified that these assessments were intended to be developmental and had no direct career consequences for the family nurses, who do not receive any financial bonus or prize

The FNP offers mothers an intensive programme of home visits, delivered in accordance with a structured curriculum. The visits incorporate content from five strands: improving mothers' physical and mental health; promoting a healthy home environment for children; supporting mothers to set and meet their own life goals (e.g. in family planning, education or employment); encouraging mothers to nurture and access their own support networks; and bolstering their knowledge of child development and confidence in parenting. As Figure A.6 shows, the last of these ('maternal role') was the single biggest programme component, accounting for around 40% of time during home visits conducted after birth. As Figure A.7 shows, visits are on average slightly longer in the months before birth, lasting just under 80 minutes; after birth, the average visit lasts just over 70 minutes. Family nurses are able to cover most of the planned content in that time; on average, they report getting through just under 95% of the content they had planned to cover.

The median mother in the FNP receives two visits per month, with 20% of mothers receiving three or more visits – equating to a median 33 visits in total (49 among those who complete the programme).[12] This is somewhat lower than the maximum of 64 visits prescribed by the FNP guidelines, but still represents very intensive support when compared with other available parenting and early years programmes in England.

Despite the highly targeted population it serves and the intensity of support it provides, the FNP operated at significant scale: at its peak in 2016, it served around 10,000 mothers in England. In that same year, the Office for National Statistics reported 22,465 live births to mothers under 20 in England and Wales - so the FNP reached around half of the target population of first-time teenage mothers in that year.

4 Empirical Strategy

4.1 Conceptual framework

A large economics literature models early human capital as the output of a production process that combines the child's endowments with parental inputs (time, goods, and behaviours) and environmental conditions (e.g. Todd and Wolpin (2003); Cunha et al. (2010); Attanasio et al. (2020)). In most empirical applications, the workforce delivering an intervention enters only implicitly through an average programme effect. We depart from this by modelling the frontline worker as an explicit input in the production of child development, adding a family nurse fixed effect to the standard human capital production function. This is motivated by the intensive, relationship-based nature of FNP home visits — which makes the identity of the assigned family nurse a plausible determinant of child outcomes.

for performance.

[12]The programme recommends weekly visits through pregnancy and until the 6th week after birth; fortnightly visits until the child is 21 months old; and monthly visits until the child is 24 months. Figure A.8 suggests that these guidelines are generally well followed, with the frequency of visits decreasing fairly continuously after birth.

We model the production of child human capital using a linear additive production function:

$$T^k_{insa} \;=\; \upsilon^k \;+\; \beta^k X_{ia} \;+\; \mu^k_n \;+\; \varepsilon^k_{insa}, \tag{1}$$

where T^k_{insa} denote child i's outcome in domain $k \in \{\text{birth health}, \text{cognitive}, \text{socio-emotional}\}$, measured at age a, for a child whose mother was visited by family nurse n in site s. X_{ia} is a vector of pre-determined child- and family-specific inputs, and μ^k_n is the FN effect on domain k; ε^k_{insa} captures unobserved inputs and measurement error. Importantly, μ^k_n captures both the direct effect the family nurse might have on children during the visit (for example via direct stimulation) and her indirect effect on the child's environment - for instance through improvements in the mother's physical and mental health, socio-economic circumstances, or parenting behaviours).

FNP nurses operate within teams that are organised into sites, and site-specific factors - such as the quality of supervision, team composition, or local services beyond the FNP - may affect children's outcomes independently of the assigned family nurse. To allow for this, we augment the production function for child human capital with a site-specific component π^k_s:

$$T^k_{insa} \;=\; \upsilon^k \;+\; \beta^k X_{ia} \;+\; \mu^k_n \;+\; \pi^k_s \;+\; \varepsilon^k_{insa} \tag{2}$$

4.2 Empirical model

Identifying μ^k_n in either of these models is challenging because mothers are rarely randomly assigned to workers. For example, mothers with particularly difficult unobserved circumstances (captured by ϵ^k_{insa}) may be assigned to particularly capable nurses (captured by μ^k_n), creating an unobserved negative correlation between μ^k_n and the error term ε^k_{insa}.

To overcome this challenge, we exploit a unique institutional feature of FNP whereby, within a team, the supervisor allocates incoming clients primarily to equalise caseload size and relative case complexity at the time of referral. Crucially, as described in section 2, this allocation occurs *before* the detailed enrolment interview and is based only on *referral-time* variables—which are known and perfectly observable by us: age, gestational age, language, and ethnicity—together with the contemporaneous distribution of caseloads across family nurses. Richer maternal characteristics (socio-economic status, health, risk behaviours, relationships) are recorded *after* allocation.

As previously described, we identify family nurse effectiveness within sites: specifically, we estimate the following equation by OLS on site-demeaned outcomes:

$$\Delta Y^k_{insat} \;\equiv\; Y^k_{insat} - \bar{Y}^k_s \;=\; \mu^k_n \;+\; \beta^k B_{it_0} \;+\; \gamma^k C_{nst_0} \;+\; \phi^k_{t_0} \;+\; \varepsilon^k_{insat}, \tag{3}$$

where Y^k_{insat} is the domain-k outcome for child i measured at age a and calendar time t; $\bar{Y}^k_s$ is the site-specific mean of that outcome; B_{it_0} are referral-time mother characteristics (age, gestational age, language, ethnicity) available to the supervisor at assignment; C_{nst_0} are the family nurse's

relative caseload size and case complexity at assignment (each standardised within site-month); and $\phi^k_{t_0}$ are month–year of enrolment fixed effects. We construct relative case complexity from a latent factor over several items collected at the enrolment interview (demographics; relationship status and contact with the biological father; socio-economic characteristics; health and health habits), then average it over the family nurse's active caseload and standardise within site-month.[13]

Consistency of our OLS estimates in equation (3) requires that $E(\varepsilon^k_{insat}|B_{it_0}, C_{nst_0}, \phi^k_{t_0}) = 0$. That is, conditional on $\{B_{it_0}, C_{nst_0}, \phi^k_{t_0}\}$, there must be no further systematic sorting on unobservables that both affect outcomes and the within-team assignment to family nurses. Two institutional features support this assumption: (i) the timing of information (the rich enrolment interview happens *after* the allocation of new clients to family nurses by the supervisor), and (ii) the explicit objective of equalising caseloads. While the assumption of quasi-random assignment is fundamentally untestable, in subsection 4.3 we present empirical evidence consistent with conditionally random mother-family nurse allocations.

Our fixed effect design will more precisely estimate effectiveness for family nurses with a larger number of clients; therefore, in our main results, we restrict our analysis sample to family nurses who have worked with at least five children with valid outcomes[14] and apply an empirical Bayes (EB) adjustment:

$$\widehat{\mu}^{k,\mathrm{EB}}_n \;=\; \lambda^k_n\,\widehat{\mu}^k_n, \qquad \lambda^k_n \;=\; \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\varepsilon/s_n}, \tag{4}$$

where s_n is the number of clients used to estimate the family nurse n's effect, σ^2_μ is the between-family nurse variance, and σ^2_ε is the residual variance. As in the teacher-effects literature (e.g., Koedel et al., 2015), shrinkage primarily trims noise at the tails.

Throughout, we implement the model by using the client's first assigned family nurse. Throughout their time in the FNP, the majority of mothers stay with the same family nurse; however, around 25% of mothers switch. In the majority of cases, this happens either because the mother moves to another location, or because the FN goes on leave, leaves the site, or leaves the FNP programme entirely. Since supervisors will have much more information about a mother already enrolled in FNP at the point of assigning her to a new FN, there is no guarantee that quasi-random assignment will hold for these pairings: we therefore restrict all of our analysis to the first family nurse each mother is assigned to.[15]

Our estimand in Equation 3 is therefore an intent-to-treat effect with respect to the initial family nurse assignment — using the first family nurse allocated to each mother, regardless of any subsequent switches. It captures the within-site contribution of the initially assigned family nurse to

[13] We define 'relative caseload size' as the family nurse's number of active clients the previous month, standardised within site-month. To account for differences in site size, we additionally control for the number of family nurses active in the site the previous month. See Appendix B for further details on the construction of this measure.

[14] In subsection 5.2, we show that the results are not sensitive to this restriction.

[15] We conduct a robustness test in which we restrict the sample to those mothers who remain with the same FN throughout in subsection 5.2.

children's outcomes, net of site-specific factors such as supervision quality and local services.

4.3 Empirical evidence of quasi-random assignment within sites

Quasi-random assignment implies that there should be no systematic correlation between a mother's large set of characteristics (E_{it_0}) collected in the post-assignment enrolment interview, and the characteristics of her assigned family nurse (Z_{inst_0}), conditional on the information set available to the supervisor at the time of assignment. We test this directly, making use of the extensive information on both mothers and family nurses collected in the FNP Information System data.

Specifically, we consider the 27 family nurse characteristics in Table 2 as well as an additional 11 measures of a family nurse's process quality (based on visits conducted before the focal mother joined the FNP): proportion of visit time spent on each of the five domains; average visit duration; average share of planned content delivered; percentage of visits where the FN gives top scores to the mother's understanding, involvement, and acceptance of the material; and average time between visits. For each of these FN characteristics, we estimate by OLS:

$$Z_{inst_0} = \alpha + \delta E_{it_0} + \beta B_{it_0} + \gamma C_{nst_{0i}} + \phi_{t_{0i}} + \pi_s + \epsilon_{inst_0} \tag{5}$$

where the vector of mother's characteristics E_{it_0} include the 19 characteristics listed in Table 1 (excluding the basic demographics, captured in B_{it_0}).

If the supervisor uses the information collected on the new client i at the time of referral exclusively for the purpose of balancing the caseload across team members and does not have access to the more detailed information in E_{it_0} at the time of assignment, we should find no systematic association between the family nurse's characteristics Z_{inst_0} and the expectant mother's characteristics E_{it_0}, conditional on the basic demographics B_{it_0}, the FN existing relative caseload size and complexity $C_{nst_{0i}}$, the site fixed effect π_s, and the year and month of enrolment $\phi_{t_{0i}}$. In other words, we should fail to reject the null hypothesis that the vector of coefficients δ are jointly zero.[16]

We perform this F-test for each of the 38 family nurse characteristics and adjust the resulting p-values for multiple hypothesis testing (MHT), given the large number of simultaneous tests and the high degree of correlation across FN characteristics. To do so, we adapt two widely used MHT algorithms - the Holm-Bonferroni-Sidak (HBS) adjustment (Holm, 1979), and the Romano-Wolf (RW) stepdown procedure (Romano and Wolf, 2005) - to the case of a joint F-test rather than a single coefficient t-test.[17]

[16] Recall that new clients are allocated to family nurses by the team supervisor, primarily to equalise caseload size and complexity across team members. As described in section 2, in 67.4% of sites there is never more than one supervisor active in the same month, and a further 14.0% have only a single month of concurrent supervisor activity consistent with a handover. For the 18.6% of sites with sustained concurrent supervisor activity, the within-site specification absorbs the average effect across any parallel supervisory structures as a site fixed effect; our estimates of individual family nurse effectiveness in these sites therefore reflect both within-team and across-team variation.

[17] We provide further detail on this adaptation in Appendix C.

In Table 3, we report the test results for the full sample at intake of mothers who joined the FNP by December 2013. For convenience, we group the 38 FN characteristics into five categories in column 1 and list the total number of outcomes in each category in column 2; we then report the number of FN characteristics for which the mothers' characteristics are jointly significant predictors when no multiple hypothesis testing correction is made (column 3), and under, respectively, the HBS and the RW algorithms (columns 4 and 5).

Without MHT correction (column 3), mothers' characteristics are significant predictors of nine of the 38 characteristics of her assigned family nurse.[18] However, after correcting for the multiplicity of the hypotheses tested, we find that none of our 38 FN outcomes are significantly predicted (at the 5% level) by mother's characteristics. Overall, the lack of a significant relationship between the client's characteristics and the FN demographics, qualifications, training, experience and process indicators strongly suggests that family nurses are, indeed, quasi-randomly assigned on these dimensions (after conditioning on the information that does influence allocation) - providing empirical support to our identification strategy.

4.4 Attrition

As in most social programmes serving disadvantaged populations, a substantial share of participants attrit from the FNP over time. Our analytical samples are therefore smaller than the full enrolment sample, and potentially selected, since the mothers retained at each stage may differ systematically from those who drop out — raising the question of whether quasi-random assignment holds within the samples used to estimate family nurse fixed effects.

By the time of the child's birth, 83% of mothers remain active in the programme; participation then falls to 66% by the time the child turns one year of age, and to 51% by the child's second birthday (when the programme ends). Attrition is somewhat higher in the first two months after enrolment but relatively stable thereafter (see Figure A.9).

Unsurprisingly, we find evidence of some selective attrition: mothers with fewer markers of disadvantage are more likely to stay in the programme for longer. Table A.1 report the means and standard deviations of selected mothers' characteristics for the enrolment sample (for women enrolled between 2007 and 2013), and for the birth and age 2 samples. However, in economic terms the differences between those who attrit and those who are retained are modest in magnitude: typically, the mothers who remain until the end of the programme are around 1-2 percentage points less likely to have markers of disadvantage such as lacking good qualifications, receiving social services

[18] These include whether the family nurse has experience working with adults as a general nurse, training in the Common Assessment Framework, or other relevant training; her average share of planned content covered in past visits; the share of past visits where she rated the mother as having top scores for understanding and for involvement; and the share of past visits dedicated to content on physical health, environmental health and the maternal life course. However, there is no consistent pattern among these correlations: for example, family nurses with longer past visits are more likely to be associated with mothers with a history of social service use, but less likely to be associated with mothers without educational qualifications (detailed results available on request). These patterns strongly suggest that these associations are indeed spurious, and corroborate the MHT-adjusted results.

or rely solely on benefits income.

Under quasi-random assignment, new clients assigned to each family nurse should - in expectation - have a similar propensity to attrit from the FNP. However, family nurses may be differentially effective at retaining mothers in the FNP. In our framework, this may cause challenges in measuring FN effectiveness: if more effective family nurses are better able to retain clients, they may end up with a more negatively selected caseload as the programme progresses. In our fixed effects framework, this will tend to bias downwards their (measured) levels of effectiveness in promoting children's development: in this case, our estimates might be best interpreted as a lower bound on true family nurse effectiveness.

We address these concerns in three ways. First, we show in Table A.2 and Table A.3 that quasi-random assignment still holds among the samples of mothers retained at different stages of the programme: in other words, we find little evidence of sorting based on FN characteristics (if anything, we see *lower* correlation between the process quality characteristics of the FNs and the mothers' characteristics at later stages of the programme). Second, we present a robustness check analysing FN effectiveness across different outcomes on a common sample of mothers, in addition to the outcome-specific maximal samples we use for our main results. As we discuss in subsection 5.2, our estimates are remarkably stable across samples, suggesting that attrition is unlikely to bias our estimates of family nurse effectiveness. Finally, we present estimates of FN effectiveness at retaining mothers into the programme in Appendix D and Table D.8. Table D.9 shows that retention effectiveness is uncorrelated with effectiveness at improving cognitive and socio-emotional development — correlations range from -0.009 to 0.073, none statistically significant at conventional levels. This provides direct evidence against the concern that more effective family nurses selectively retain harder-to-reach mothers, which would otherwise bias our estimates of FN effectiveness at improving child outcomes downward.[19]

4.5 Family nurse versus site-specific effects

Our preferred specification recovers the distribution of family nurse fixed effects within sites and therefore abstracts from site-specific factors. An interesting question, however, relates to the relative contribution of family nurse vs site-specific factors to children's outcomes. To be able to answer this question, one would need to separately identify family nurse from site fixed effects, which would require that a) enough family nurses genuinely switch sites and b) the quasi-random allocation of clients to family nurses holds across sites (as well as within).

These requirements are stringent for our context because a) most family nurses are observed at a single site and genuine "switchers" are rare, and b) in principle, the allocation process takes place within sites. In practice, however, we find that family nurses do appear to be quasi-randomly allocated to mothers even *across* sites: when we re-estimate Equation 5 without site fixed effects,

[19]Section D also reports the distribution of family nurse effectiveness at retaining mothers in the programme (Figure D.21) and its correlates with structural and process quality indicators (Figure D.22).

we find that - after adjusting for multiple hypothesis testing using the Holm-Bonferroni-Sidak algorithm - mother characteristics are significant predictors of 2 out of 38 FN characteristics, or 5.3%.[20]

In light of this, we also present estimates of the distribution of family nurse fixed effects in a specification with no site fixed effects:

$$Y^k_{insat} = \mu^k_n + \beta^k B_{it_0} + \gamma^k C_{nst_0} + \phi^k_{t_0} + \varepsilon^k_{insat}, \tag{6}$$

In this specification, the parameters μ^k_n have a different interpretation from those estimated in equation (3) in that, if the underlying model does include site-specific effects on children's outcomes, the parameters μ^k_n will capture both family nurse and site-specific effects (and any interactions between them). While these estimates should be seen as more tentative than the estimates of FN fixed effects within sites, they are interesting to compare to the distribution of FN effects from the within site specification to get a sense of how important site-specific factors (and their possible interactions with family nurse-specific factors) might be relative to family nurse-specific factors. We discuss them after discussing the within-site results.

4.6 Measurement of key outcomes

We estimate Equation 3 above for five outcomes, which measure the development of children in different dimensions and at different ages. This includes a measure of child's **health at birth**, which is constructed as a factor score measured by the child's gestational age, birth weight and the number of weeks spent in neonatal intensive care.[21] The other four children's outcomes are two measures of cognitive skills and two measures of socio-emotional development, each based on the Ages and Stages Questionnaire (ASQ) and collected both at age 1 and at age 2.[22] The measures of **cognitive development** are based on the ASQ-3 and cover five domains: communication, gross motor skills, fine motor skills, problem solving, and personal social skills. The measures of **socio-emotional development** are based on the ASQ-Socio Emotional (ASQ-SE). The ASQ-3 and ASQ-SE are well validated developmental screeners for cognitive and socio-emotional dimensions, with strong predictive power for outcomes at later ages (Schonhaut et al., 2020, 2021).[23]

[20] Without adjusting for MHT, mother characteristics significantly predict 4 nurse characteristics. Results are available in Table A.4 for the full sample, and in Table A.5 and Table A.6 for outcome-specific samples.

[21] We construct a factor score in order to combine different but related measures to extract information about the infant's underlying health status. While these measures are self-reported by the mothers, they reflect clinical facts recorded at birth that are known precisely from hospital records.

[22] There is considerable variation in the precise age at which the ASQ 12- and 24-month assessments were administered, so we always work with scores residualised on age (months at the time of the test) and gender (to account for differences in developmental velocity between boys and girls). Family nurses were instructed to use the version of the ASQ most appropriate to the child's actual age. For our 12-month measure, we therefore pool the results for children assessed using the 10-month and the 14-month questionnaire; where results are available for both questionnaires, we take a simple average of the two (residualised) scores. We conduct a similar procedure for the 24-month measure, in that case pooling results from the 20-month and 24-month questionnaires.

[23] Kendall et al. (2019) shows that the ASQ-3 is both acceptable and understandable as measure of child development at age 2–2.5 years by parents and health professionals in England; it is now included as part of England's universal

Both the ASQ-3 and ASQ-SE have scores ranging from 0 to 60, where a higher score indicates better (worse) cognitive (socio-emotional) development.[24] To construct our outcome variables, we reverse the scoring of the ASQ-SE in our analysis so that a higher score indicates better socio-emotional development. We then winsorise the distributions at the first percentile and we standardise them to have mean 0 and standard deviation 1.

5 Estimates of family nurse effectiveness

5.1 Distribution of measures of family nurse effectiveness

Panel A of Table 4 reports various moments of the estimated distributions of the family nurse fixed effects for each of our five child outcomes of interest, using within-site variation. Figure 1 plots the corresponding distributions.

Across outcomes, we find meaningful heterogeneity in family nurse effectiveness. The distribution is narrowest for children's birth outcomes, when mothers have had less exposure to their family nurse (and interestingly where the RCT found least evidence that the FNP was an effective intervention, see Robling et al. (2015)). We find that a one-standard deviation (SD) increase in family nurse effectiveness leads to a 0.109 SD improvement in our birth outcome factor.

Distributions of family nurse effectiveness are wider for children's developmental outcomes. For cognitive outcomes, a one-SD increase in family nurse effectiveness improves cognition by 0.22SD at age 1 and by 0.20SD at age 2, respectively. We find the widest distribution for socio-emotional outcomes: here, a one-SD increase in family nurse effectiveness improves scores on the ASQ-SE by 0.31 SD at age 1, and by 0.23SD at age 2, respectively.

As Figure 1 shows, the wider distributions are driven by a much longer left tail, meaning that variation comes more from family nurses who significantly under-perform the average FN in their site. Importantly, since these outcomes are measured *relative* to the average family nurse within a site, we cannot conclude from our analysis whether less effective FNs are actively *harmful* (i.e., children would have had better development without the FNP), or whether they simply deliver (much) smaller benefits for child development than their more effective peers.

Results in Panel B of Table 4 report statistics of the distribution of across-site FN fixed effects. Their standard deviation is consistently 20-25% larger than the within-site distribution in Panel A: this suggests that site-level factors, such as supervisor characteristics or site management (or other local services), are somewhat important in driving overall variation in the outcomes of children across

Age 2 to 2-and-a-half health checks carried out by health visiting teams (Conti et al., 2026).

[24] Rather than using the full distribution of the ASQ and ASQ-SE, many previous studies have constructed categorical variables indicating whether the child is in the 'referral zone' for possible developmental problems. However, these referral zones are identified using cut-offs which are specific to the US population; no such values exist for the UK context. Additionally, continuous outcomes tend to perform better in models with a large number of fixed effects (our measures of FN effectiveness). We therefore follow a large literature, including Doyle et al. (2013), in using the full domain of ASQ scores and modelling them as continuous variables.

family nurses and sites, but there remains considerable variation that is linked to the individual family nurse. For this reason and because our identification strategy is stronger for the within-site distribution of FN effectiveness, we focus on the within-site measures of family nurse effectiveness in the remainder of the paper.

Randomisation inference In order to assess the statistical significance of these results, it is important to have a sense of how wide the distribution of family nurse effectiveness is, as compared to what would have been expected by chance. Hence, we adopt a randomisation inference approach to assess whether our estimates suggest an 'unexpected' degree of heterogeneity in family nurse effectiveness. To do this, we first randomly reshuffle family nurses across mothers and children (within sites). We then estimate a set of FN fixed effects based on this randomised dataset, and compute the standard deviation of this distribution as a summary statistic for the amount of variation in family nurse fixed effects when the true effect is known to be zero. We repeat the exercise 500 times for each of our five outcomes. In Figure 2, we plot (in beige) the distribution of the standard deviation of the FN fixed effects arising from these 500 permutations: the red vertical lines show the 5th and 95th percentiles of this distribution. The vertical green line shows the estimated standard deviation based on the actual FN allocation, as reported in Table 4. Where the standard deviation of the true FN effectiveness distribution falls within the bounds of our randomised distribution, we cannot rule out estimating a similarly wide distribution even in the absence of any actual differences in FN effectiveness.[25]

Instead, for each of our five outcomes, the standard deviation of the family nurse effectiveness based on the assigned family nurse lies outside 90% of the permutation distribution, indicating that the true degree of heterogeneity is larger than we would expect to observe by chance. Indeed, for our cognitive and socio-emotional outcomes, the true standard deviation of FN effectiveness lies well outside the distribution of our randomisation inference test. This means that there is substantially more heterogeneity than we would expect by chance: for example, the standard deviation of family nurse effectiveness for age 2 socio-emotional outcomes is twice as large as the median of the permutation distribution. Only measured FN effectiveness at improving birth outcomes lies closer to the randomisation inference distribution: this indicates that, although there is a statistically significant degree of heterogeneity, its magnitude is still small relative to the heterogeneity present due to pure noise.

5.2 Robustness and sensitivity checks

Robustness to failures of quasi-random assignment As we discussed in section 4, our main identification assumption is that - conditional on a limited set of observable characteristics and

[25] The non-zero heterogeneity in effectiveness in our randomisation inference tests essentially reflects the fact that family nurses have a finite caseload. While FN characteristics are, on average, unrelated to family characteristics, some family nurses will randomly end up with a caseload of clients whose children would have had better-than-average outcomes regardless of the family nurse assigned to them: this will affect both the child's outcomes and the measured FN effectiveness.

fixed effects - new mothers joining the FNP are randomly assigned to a family nurse. In our main specifications, we therefore account for site-specific factors by demeaning outcomes within sites, and additionally control for basic data collected on a mother's referral form, month-year of enrolment fixed effects, and the family nurse's relative caseload size and complexity.

In Table 5 we test whether our estimates of family nurse effectiveness are robust to changes in the control set: both dropping our standard controls for caseload size and complexity and mother demographics, and including a range of additional information about the mother.[26]

As shown in Table 5, the results are virtually identical when we remove the caseload controls (panel B) or when we remove those for the size and complexity of the caseload and the basic demographics of the mother (panel C).[27] This is also reassuring evidence that, even if our proxy for caseload size and complexity imperfectly captures the way supervisors assess and take into account existing caseloads, our results would be robust to this misspecification. As shown in the table and in Figure A.10, results are also robust to the inclusion of mothers' extended characteristics at enrolment as controls (Panel D). This provides further evidence consistent with our identification assumption of quasi-random assignment holding in practice.

Sensitivity to sample selection In Panel E of Table 5 and in Figure A.11, we show the impact of imposing a common sample across our outcomes of interest. This common sample uses only children with valid results for all outcomes to estimate FN effectiveness, with a consequent loss of sample as compared to the benchmark results when we include all the available data points (Panel E). Imposing this common sample leads to a slightly wider distribution of FN effectiveness for the birth outcomes, for which the drop in sample size is bigger; the results for all other outcomes look similar to our main estimates.

We also assess the sensitivity of our results to the restriction on the minimum number of clients per nurse. As mentioned above, our analysis sample of family nurses is restricted to those with at least five children with valid outcomes. In Table 5 (Panels F and G) and Figure A.12, we analyse how sensitive our estimates of FN effectiveness are to this restriction. We find reassuringly little difference between the various specifications. To the extent that there are differences, we find that sample restrictions with a lower threshold for the minimum number of clients (Panel F) yield slightly larger estimates of the variation in FN effectiveness than our main results (Panel A), while restrictions that impose a higher threshold yield a tighter distribution (Panel G). However, these differences are very small, leading us to conclude that our benchmark threshold of five or more mothers per family nurse strikes the right balance between improving precision and avoiding unnecessary sample loss.

[26] The mother characteristics are the set of controls E_{it_0} which are collected at enrolment (and which we use in the test of quasi-random assignment). To the extent that our random assignment assumption is violated in ways that correlate with any of these variables, we would expect their inclusion to shift our estimates of FN effectiveness.

[27] Recall that these controls are included because they were known to FNP supervisors at the time of assignment, and so in principle could have affected it. Mother demographics include age (in months) at enrolment; gestational age at enrolment; whether the mother's primary language is English; and mother ethnicity.

Finally, we assess sensitivity to the inclusion of sites with more complex supervisory structures. As discussed in footnote 4, 18.6% of sites have two or more months with simultaneously active supervisors, consistent with job-share arrangements or the operation of multiple teams within larger sites. In Panel H of Table 5, we re-estimate our main specification excluding these 24 sites, and find that the results are reassuringly similar. Hence, our main results are not driven by sites with more complex supervisory structures.

Measurement of child development Child development at ages 1 and 2 is measured using the ASQ-3 and ASQ-SE, which are parent-reported instruments administered with the support of the family nurse. A potential concern is that more effective family nurses may also be better at eliciting responses from mothers — for example, by explaining items more carefully or encouraging more accurate reporting — such that part of the estimated dispersion in family nurse effectiveness could reflect differential measurement quality rather than differential impact on child development. While we do not have an objective measure of child development to test this directly, several features of our results weigh against this interpretation. First, as we show in Section 6, the pattern of comparative advantage across developmental domains, with within-domain correlations across ages substantially stronger than across-domain correlations, is difficult to reconcile with a pure measurement story, which would tend to inflate correlations across all outcomes uniformly. Second, as we show in Section 7, the mechanisms through which effective family nurses operate reflect substantive changes in maternal behaviour that would not follow from differential reporting alone.

Leave-out estimates Our main estimates calculate each family nurse's effectiveness based on the outcomes of all of her assigned clients within our analysis sample. An alternative approach is to adopt a 'leave-out' approach to calculate measures of family nurse effectiveness, and to use these to perform the forecast bias test proposed by Chetty et al. (2014a). For each mother i, we re-estimate the full set of family nurse fixed effects on the sample excluding i, yielding a measure $\hat{\mu}_{n,-i}$ that is mechanically independent of mother i's outcome. We repeat this procedure for all clients and all outcomes in our analysis sample. We then regress outcomes on these leave-out measures, progressively adding controls, time fixed effects, and site fixed effects. Under no forecast bias in a model with homogeneous nurse effects, the coefficient on the leave-out estimate should equal one.

Appendix Table A.7 reports the results. For the cognitive and socio-emotional outcomes at ages 1 and 2, the coefficients range from 0.73 to 0.91 across specifications — large, stable as controls are added, and consistent with the leave-out measures being strong predictors of child outcomes. Coefficients may fall below one for several reasons. First, our empirical Bayes adjustment will compress the value of the regressor, yielding a coefficient less than one. Second, there may be additional richness in the model that is not captured in our estimates of family nurse effectiveness, including changes in FN effectiveness over time or match quality effects that aren't captured in a model with homogeneous FN effects. For birth outcomes, the coefficients are smaller (0.12–0.19), in line with the weaker family nurse effects on this outcome documented above.

6 What characterises relatively more effective family nurses?

So far, we have shown that there is a substantial degree of heterogeneity in family nurses' effectiveness at improving child development, particularly with respect to cognitive and socio-emotional development. This means that - even within a standardised programme delivered by a highly qualified, highly trained, highly paid workforce such as the FNP - there are substantial developmental returns to being assigned a more effective family nurse. In this section, we aim to better understand the characteristics and behaviours that are associated with more effective family nurses and the mechanisms through which they seem to promote better outcomes in the children they support.

6.1 Correlation between different domains of effectiveness

Are some family nurses particularly strong at promoting one dimension of child development relative to others, or does higher effectiveness applies to all domains of child development? To shed light on this question, Table 6 reports the correlation between our estimates of family nurse effectiveness along different dimensions. Across all outcomes, measures of family nurse effectiveness are positively correlated with each other, but two patterns are of note.

First, the correlations are much stronger within the cognitive and socio-emotional domains across ages than across these domains within ages. For example, there is a 67% correlation between a family nurse's effectiveness at promoting cognitive development at age 1 and at age 2, but the correlation in family nurse effectiveness across cognitive and socio-emotional outcomes is around 25-37%. This suggests that comparative advantage rather than absolute advantage might best describe this workforce: while family nurses who are good at improving one domain tend to be good at improving another domain, they have particular specialisms or areas of relative strength.

Second, we find much weaker correlations, if any, between family nurse effectiveness at promoting better health at birth and better cognitive development, on the one hand, and no significant relationship between effectiveness at promoting birth outcomes and subsequent socio-emotional development, on the other. This could suggest that the traits or behaviours that make family nurses relatively more effective at improving cognitive development are more similar to those than make them effective at improving socio-emotional development than at improving child health at birth. It could also reflect the much smaller degree of true variation in FN effectiveness at promoting health at birth, which makes the ordinal ranking of family nurse effectiveness in this dimension more uncertain.

6.2 Predictors of variation in family nurse effectiveness

Building on this insight, we next study the traits and behaviours that predict family nurse effectiveness across domains of development. To do so, we follow the literature on early years workforce quality and distinguish between markers of 'structural quality' and of 'process quality'. *Structural quality* refers to the family nurse characteristics that are objective, observable information and that

would, in most cases, be commonly available to an employer during a hiring process: these include demographic characteristics, experience, qualifications and training. We add to these standard measures two predictors of 'match quality': the proportion of a FN caseload who matches her ethnicity[28] (ethnic minority or not), and the proportion of a FN caseload who matches her experience (clients aged 17 or younger for FNs with experience working with teen parents, or 18 and above for FNs without): to the extent that these markers are observable during the FN hiring process, they could help the FNP to hire more workers with relevant attributes.

Measures of *'process quality'*, instead, capture what family nurses actually *do* with the mother-child dyad during the visits, including the quality of their interactions. These measures are typically more difficult to assess during a hiring process. Even so, a better understanding of how process quality relates to effectiveness would support better decision-making by FNP managers: for example, it could inform decisions on how to prioritise retention efforts, the allocation of additional training and support, or the assignment of more effective family nurses to families with greater levels of disadvantage.

To measure process quality in our context, we exploit two sources of information. The first is the assessments carried out by supervisors accompanying family nurses in their home visits: as outlined in section 3, they score the visits they attend according to 44 different criteria, aggregated under 10 categories (e.g. planning and scheduling visits; reviewing the previous visit; setting goals; or delivering programme content). We first verify, using an event study design, that supervisory visits did not themselves affect the quality or content of home visits as reported by the family nurse (Figure A.2 and Figure A.3), supporting their use as independent quality indicators.

The second source of process quality indicators comes from the metadata on each visit conducted by a family nurse (accompanied or not); this covers: the length of the visit, the proportion of planned content covered, and the proportion of time spent on content from each of the five domains of the FNP programme. We show in Figure A.13 that there is some variation across family nurses along these dimensions, particularly in the average visit duration. Our measures of process quality also include the family nurse's assessment at the end of each visit of how engaged the mother was, how well the mother understood the material, and the extent to which the mother showed 'conflict' with it. We view these self-assessments as containing information not only about the mother, but also about her family nurse, since building a trusting relationship with mothers and presenting the material in a way that they feel engaged with is a core part of the role.

In order to understand whether and to which extent these structural and process quality indicators predict family nurse effectiveness, we first assess their pairwise correlations, and then we test the joint significance of groups of indicators (demographics, family nurse type, qualifications, experience and training, FN-mother match, supervisor assessment, visit metadata, programme content). We comment here on the correlations of structural and process quality indicators with the FN

[28] Recall that client's ethnicity is collected in the referral form, and that supervisors seek to match clients with family nurses by ethnicity.

effectiveness at improving child cognitive development at age 2, displayed in Figure 3. Results for cognitive development at age 1 and socio-emotional development at ages 1 and 2 are shown in Figure A.14, Figure A.15 and Figure A.16, respectively, and broadly show qualitatively similar patterns.

We first comment on the predictive power of structural quality indicators for FN effectiveness across different domains. Panel (a) of Figure 3 shows little association between a family nurse's observable structural quality characteristics and her effectiveness in promoting child cognitive development at age 2: family nurses with experience working as a school nurse tend to be somewhat more effective, while those who have done some post-graduate study (without completing it) tend to be somewhat less effective; however, in all cases these correlations are relatively small. Additionally, matching a family nurse and a mother based on their specialist experience is negatively correlated with the FN effectiveness.

In panel (b) of Figure 3, we continue this exercise by looking at the correlations between family nurse effectiveness and her average process quality measures based on supervisor assessments,[29] visit metadata and programme content (the latter two sets as reported by the FN herself). Despite the predictive power of independent assessments of process quality in other contexts, we find that supervisors' assessments hold little information about family nurse effectiveness: the ten assessment domains are not jointly significant predictors, and none of them individually correlates significantly with FN effectiveness.[30] We further find no relationship between a FN effectiveness and the characteristics of her supervisor or team-mates (Figure A.18).[31]

By contrast, the visit metadata and programme content data recorded after each visit are more predictive of effectiveness: family nurses who report having longer visits, covering more content, seeing better engagement and understanding from the mothers, and spending relatively more (less) time on the home environment (maternal role) domain are, on average, significantly more effective at boosting child development. That these visit metadata indicators are more predictive of family nurse effectiveness than structural quality measures is in some sense unsurprising, since the process quality indicators will capture information about the implementation of the programme and the ongoing relationship between family nurse and mother. These results are nonetheless important, since they highlight the critical role of programme implementation in delivering successful outcomes, and suggest that relatively low-cost reporting can be an effective tool for tracking programme implementation.

These measures of a family nurse's process quality are computed across her full caseload, and therefore include the clients used to estimate her effectiveness. A potential concern is that me-

[29] To reduce dimensionality, we aggregate the 44 different supervisor scores by taking the total score for each of the 10 sub-categories of the assessment.

[30] This may partly reflect the very high scores given by supervisors: on average, family nurses receive a score of 94% during the supervisory visit; across all individual questions, supervisors give out the top rating in 72% of cases.

[31] We use the characteristics of the FN modal supervisor, and of the family nurses who overlap at the same site as the focal one, weighted by the number of months of overlap.

chanical correlation could arise if mothers who are predisposed to engage generate both longer visits and better child outcomes independently of any FN quality. To address this, Figure A.17 redoes the analysis at the client level, correlating each client's process quality measures with a leave-out measure of her assigned FN's effectiveness — constructed excluding that client from the effectiveness estimation, following the approach in Section 5.2. This design removes any mechanical link between a given mother's outcomes and her family nurse's measured effectiveness. The results are qualitatively similar to the FN-level averages presented here, though somewhat more statistically significant. This likely reflects the greater variation in process quality measures across individual client experiences compared to FN-level averages, where collapsing to a single mean per FN compresses much of the relevant signal. Together, these results suggest that the mechanical correlation concern is unlikely to be driving our main finding that process quality predicts family nurse effectiveness whereas structural quality does not.

7 How do effective family nurses support child development?

By design, the FNP programme aims to influence several inputs in the production of child health and development: indeed, content aimed at directly improving parenting - introducing stimulating activities, building connections between mother and child, helping mothers to understand the process of child development - makes up around 40% of the visits following birth (see Figure A.6). To achieve these goals, the FNP programme also aims to improve mothers' capacity to undertake these positive parenting behaviours, in different ways: supporting the mother's mental health; trying to reduce risky behaviours, to improve her social and economic circumstances (known as 'life course' within the FNP) and to develop strong, supportive relationships; and encouraging appropriate investments into the child's health and development. In common with most home visiting-based parenting programmes, the FNP theory of change emphasises the central role of parents as the child's primary carers, rather than the direct influence of the family nurse herself during the home visits.

Empirical approach In this section, we explore the channels through which effective family nurses operate, by exploiting rich data collected throughout the FNP programme on mothers' risky behaviours, socio-economic circumstances, relationships and mental health, as well as wider investments in the child. For each of these five possible mechanisms, we use factor analysis to construct summary indices at the child's ages 12 and 24 months; we provide further detail on the construction of each of these measures in Appendix E. We then run simple linear regressions of the potential mechanisms on the relevant measure of family nurse effectiveness. Note that we focus on the child's outcomes at ages 1 and 2 because we lack measures of most of our potential mechanisms collected during pregnancy (but after intake).

There is potential for confounding in this exercise: while mothers are randomly assigned to family nurses at the time of enrolment, sampling variation in the finite samples of mothers per family nurse

mean that some FNs will randomly be allocated a 'better' caseload. These groups may have both better-than-average mother characteristics and so better-than-average child outcomes, which in turn will lead to higher measured family nurse effectiveness. This means that a simple comparison between overall measured FN effectiveness and maternal behaviours may not recover a true causal effect of the family nurse.

In keeping with the literature (e.g. Bacher-Hicks and Koedel, 2023), we therefore use the 'leave-out' measures of family nurse effectiveness described in subsection 5.2, which for each client i are estimated on all mothers except i herself. This removes any mechanical correlation between a mother's own circumstances and behaviours, and our measure of the effectiveness of her family nurse.

Results In Table 7, we show the impact of family nurse effectiveness on each of the five potential mechanisms: maternal risky behaviours; maternal socio-economic circumstances; maternal relationships; maternal mental health; and investments in child development. As these five variables are correlated with each other, we report both the unadjusted p-value (below the standard error) as well as the p-value adjusted for multiple hypothesis testing using the Romano-Wolf (RW) stepdown procedure (Romano and Wolf, 2005).

We find that mothers assigned to family nurses who are more effective at improving children's cognitive development both at age 1 and age 2 are also significantly less likely to display risky behaviours like smoking, drinking, drug use, or having unprotected intercourse. Mothers whose family nurses better promote cognitive development at age 1 also develop more favourable socio-economic circumstances (including education and employment), are more likely to have positive relationships (for example, greater contact with their child's father or less exposure to abuse), and have better mental health, but none of these relationships remain statistically significant once we adjust for multiple hypotheses. At age 2, investments in children (including vaccinations, breastfeeding, and use of wider early years services) are also strongly correlated with having a family nurse who is more effective at promoting cognitive development. Since these FN effectiveness measures are calculated on all families *except* the focal mother, these results indeed suggest that more effective family nurses are causally improving some of these behaviours.

We see quite a different pattern when focusing on family nurses who are particularly good at promoting children's socio-emotional development: they consistently seem to be particularly effective at improving mothers' mental health (at ages 1 and 2).

8 Conclusions

There is a large and rich evidence base on the impacts of early years interventions on children's outcomes throughout their lives. However, there is now also a large volume of research finding that some interventions fail to live up to their potential, particularly when they are delivered at

scale. Understanding what drives this variation in effectiveness is crucial for policy design and implementation.

Our paper provides the first large-scale evidence on the importance of workforce quality in delivering effective early childhood interventions via home visiting. Exploiting the quasi-random assignment of mothers to family nurses in England's Family Nurse Partnership programme, and combining this with exceptionally rich data on both workers and families, we document three key findings that advance our understanding of how workforce quality shapes programme effectiveness.

First, there is substantial variation in worker effectiveness, even within a highly structured programme employing qualified professionals. A one-standard deviation increase in family nurse quality leads to a 0.20 SD improvement in cognitive development and a 0.23 SD improvement in socio-emotional development at child age 2. This magnitude of impact - approximately twice what would be expected by chance - demonstrates that individual worker quality is a crucial determinant of programme success. These effects are particularly notable given the programme's highly structured curriculum and the workforce's high level of qualification: all family nurses are registered nurses or midwives, educated to degree level, undergo extensive additional training, and are paid more than standard nursing roles. The substantial heterogeneity in effectiveness even in this context suggests that workforce quality issues may be even more pronounced in less structured programmes or those with less qualified staff.

Second, despite having access to an unusually rich set of potential predictors of family nurse effectiveness, we find that traditional 'structural' quality measures have minimal predictive power. Observable characteristics observable at hiring, including education and experience, are barely predictive of effectiveness. Perhaps surprisingly, even detailed supervisory assessments conducted during accompanied visits show little relationship to measured FN effectiveness. The strongest predictors of quality come from the visit implementation data collected by the family nurse herself - including both objective measures like visit duration and more subjective assessments of maternal engagement. This finding is particularly important, as it suggests that relatively low-cost administrative data could help programme managers identify implementation challenges and target additional support where needed. The difficulty in predicting worker effectiveness from traditional measures also raises important questions about hiring practices and professional development in early years services.

Third, we identify key mechanisms through which effective family nurses operate, providing new insights into how high-quality early interventions actually achieve their impacts. Family nurses who excel at promoting cognitive development appear to be effective at reducing maternal risk behaviours during pregnancy and the early years, while those most successful at improving children's socio-emotional outcomes appear to be effective at improving maternal mental health. These distinct pathways suggest that effective family nurses are not simply 'better at everything' but rather succeed through specific channels. The correlation between effectiveness at improving different domains of child development is positive but modest, suggesting that family nurses may have com-

parative advantages in supporting different aspects of child development.

These findings have important implications for scaling up early childhood programmes. While standardised curricula and workforce professionalisation are commonly proposed solutions for maintaining programme effectiveness at scale, our results suggest that these alone may be insufficient to ensure consistent quality: high levels of qualification and extensive training do not appear to eliminate substantial variation in effectiveness across workers. At the same time, our finding that implementation quality - how the programme is delivered - matters more than structural factors like formal qualifications suggests promising avenues for improvement: process data monitoring could provide a valuable tool for identifying implementation challenges and targeting support.

More broadly, our findings contribute to ongoing debates about how to maintain programme effectiveness at scale. Previous research has made significant progress in understanding 'what works' through rigorous evidence and replication, as well as 'for whom' through robust analyses of targeting and heterogeneous effects. However, much less is known about the 'why' - particularly on the supply side. Our results demonstrate that workforce quality is a crucial factor in maintaining programme effectiveness at scale, even in highly structured interventions with well-qualified staff. Crucially, we show that it is not observable credentials or qualifications that distinguish more from less effective workers, but rather how the programme is actually delivered — the quality of interactions, the engagement of clients, and the content prioritised during visits. This points to a concrete and actionable lesson for successful scaling: monitoring implementation quality through routinely-collected low-cost administrative data, and targeting training and support at the process dimensions that predict effectiveness, might be more promising levers than hiring on the basis of formal credentials alone.

References

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Andrew, A., O. P. Attanasio, R. Bernal, L. C. Sosa, S. Krutikova, and M. Rubio-Codina (2024). Preschool quality and child development. *Journal of Political Economy 132*(7), 000–000.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016). Teacher Quality and Learning Outcomes in Kindergarten. *The Quarterly Journal of Economics 131*(3), 1415–1453.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. NBER Working Paper W20965.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

Bacher-Hicks, A. and C. Koedel (2023). Estimation and interpretation of teacher value added in research applications. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 6 of *Handbook of the Economics of Education*, pp. 93–134. North-Holland.

Bald, A., J. J. Doyle Jr, M. Gross, and B. A. Jacob (2022). Economics of foster care. *Journal of Economic Perspectives 36*(2), 223–246.

Barnes, J., M. Ball, P. Meadows, J. McLeish, and J. Belsky (2008). Nurse-family partnership programme: First year pilot sites implementation in england. *London: Institute for the Study of Children, Families and Social Issues, Birkbeck, University of London*.

Blanden, J., K. Hansen, and S. McNally (2017, February). Quality in Early Years Settings and Children's School Achievement. CEP Discussion Papers dp1468, Centre for Economic Performance, LSE.

Blau, D. and J. Currie (2006). Pre-school, day care, and after-school care: who's minding the kids? *Handbook of the Economics of Education 2*, 1163–1278.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a, September). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b, September). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–79.

Conti, G., S. Kliem, and M. Sandner (2024). Early home visiting delivery model and maternal and child mental health at primary school age. In *AEA Papers and Proceedings*, Volume 114, pp. 401–406. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Conti, G., M. Sandner, T. Brand, and S. Kliem (2025). Efficacy of delivery models in early intervention: Findings from germany's nurse-family partnership on family and child welfare services and pediatric medical incidents in high-risk families. *Child Abuse & Neglect 166*, 107513.

Conti, G., J. Smith, E. Anson, S. Groth, M. Knudtson, A. Salvati, and D. Olds (2024). Early home visits and health outcomes in low-income mothers and offspring: 18-year follow-up of a randomized clinical trial. *JAMA network open 7*(1), e2351752.

Conti, G., G. Weber Costa, D. D'Mello, and Y. Yu (2026). The state of health visiting in england: Workforce composition, caseloads and service delivery. *medRxiv*, 2026–03.

Cunha, F., J. Heckman, and S. Schennach (2010). Estimating the technology of cognitive and non-cognitive skill formation. *Econometrica 78*(3), 883–931.

Currie, J. and J. Zhang (2024). Doing more with less: Predicting primary care provider effectiveness. *Review of Economics and Statistics*, 1–16.

Dahlstrand, A. (2025). *Defying Distance?: The Provision of Medical Services in the Digital Age.*

Doyle, O., C. Harmon, J. Heckman, C. Logue, and S. H. Moon (2013). Measuring investment in human capital formation: An experimental analysis of early life outcomes. Working Paper Series, UCD Centre for Economic Research, No. 13/10.

Doyle Jr, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review 97*(5), 1583–1610.

Duncan, G., A. Kalil, M. Mogstad, and M. Rege (2023). Investing in early childhood development in preschool and at home. *Handbook of the Economics of Education 6*, 1–91.

Eckenrode, J., M. Campa, D. W. Luckey, C. R. J. Henderson, R. Cole, H. Kitzman, E. Anson, K. Sidora-Arcoleo, J. Powers, and D. L. Olds (2010). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatrics & Adolescent Medicine 164*(1), 9–15.

FNP National Unit (2017, August). Fnp eligibility criteria, managing client allocation, recruitment and enrolment. Clinical guidance, Tavistock and Portman NHS Foundation Trust. For use by FNP licensed sites only.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2024). Augmenting state capacity for child development: Experimental evidence from india. *Journal of Political Economy 132*(5), 1565–1602.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jacob, B. A., L. Lefgren, and D. P. Sims (2010). The persistence of teacher-induced learning. *Journal of Human Resources 45*(5), 915–943.

Kendall, S., A. Nash, A. Braun, G. Bastug, E. Rougeaux, and H. Bedford (2019). Acceptability and understanding of the ages & stages questionnaires®, third edition, as part of the healthy child programme 2-year health and development review in england: Parent and professional perspectives. *Child: Care, Health and Development 45*(2), 251–256.

Kitzman, H., D. L. Olds, C. R. J. Henderson, C. Hanks, R. Cole, R. Tatelbaum, K. M. McConnochie, K. Sidora, D. W. Luckey, D. Shaver, K. Engelhardt, D. James, and K. Barnard (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: A randomized controlled trial. *JAMA 278*(8), 644–652.

Kitzman, H., D. L. Olds, M. D. Knudtson, R. Cole, E. Anson, J. A. Smith, D. Fishbein, R. DiClemente, G. Wingood, A. M. Caliendo, and G. Conti (2019). Prenatal and infancy nurse home visiting and 18-year outcomes of a randomized trial. *Pediatrics 144*(6), e20183876.

Kitzman, H., D. L. Olds, K. Sidora, C. R. J. Henderson, C. Hanks, R. Cole, D. W. Luckey, J. Bondy, K. Cole, and J. Glazner (2000). Enduring effects of nurse home visitation on maternal life course: a 3-year follow-up of a randomized trial. *JAMA 283*(15), 1983–1989.

Kitzman, H. J., D. L. Olds, R. E. Cole, C. A. Hanks, E. A. Anson, K. J. Arcoleo, D. W. Luckey, M. D. Knudtson, C. R. J. Henderson, and J. R. Holmberg (2010). Enduring effects of prenatal and infancy home visiting by nurses on children: Age-12 follow-up of a randomized trial. *Archives of Pediatrics & Adolescent Medicine 164*(5), 412–418.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015). Value-added modeling: A review. *Economics of Education Review 47*(C), 180–195.

Leer, J. and F. Lopez-Boo (2019). Assessing the quality of home visit parenting programs in latin america and the caribbean. *Early Child Development and Care 189*(13), 2183–2196.

List, J. A. (2026). The economics of scaling early childhood programs: Lessons from the chicago school. Technical report, National Bureau of Economic Research.

List, J. A., D. Suskind, and L. H. Supplee (2021). The scale-up effect in early childhood and public policy. *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It Routledge*.

McConnell, M. A., S. Rokicki, S. Ayers, F. Allouch, N. Perreault, R. A. Gourevitch, M. W. Martin, R. A. Zhou, C. Zera, M. R. Hacker, et al. (2022). Effect of an intensive nurse home visiting program on adverse birth outcomes in a medicaid-eligible population: a randomized clinical trial. *Jama 328*(1), 27–37.

Olds, D., C. R. J. Henderson, R. Cole, J. Eckenrode, H. Kitzman, D. Luckey, L. Pettitt, K. Sidora, P. Morris, and J. Powers (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA 280*(14), 1238–1244.

Olds, D. L., J. Eckenrode, C. R. J. Henderson, H. Kitzman, J. Powers, R. Cole, K. Sidora, P. Morris, L. M. Pettitt, and D. Luckey (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year follow-up of a randomized trial. *JAMA 278*(8), 637–643.

Olds, D. L., C. R. J. Henderson, R. Chamberlin, and R. Tatelbaum (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics 78*, 65–78.

Olds, D. L., C. R. J. Henderson, R. Tatelbaum, and R. Chamberlin (1986). Improving the delivery of prenatal care and outcomes of pregnancy: A randomized trial of nurse home visitation. *Pediatrics 77*(1), 16–28.

Olds, D. L., J. R. Holmberg, N. Donelan-McCall, D. W. Luckey, M. D. Knudtson, J. Robinson, Y. Xie, and L. M. Pettitt (2014). Effects of home visits by paraprofessionals and by nurses on children: Age-six and nine follow-up of a randomized trial. *JAMA Pediatrics 168*(2), 114–121.

Olds, D. L., H. Kitzman, E. Anson, J. A. Smith, M. D. Knudtson, T. Miller, R. Cole, C. Hopfer, and G. Conti (2019). Prenatal and infancy nurse home visiting effects on mothers: 18-year follow-up of a randomized trial. *Pediatrics 144*(6), e20183889.

Olds, D. L., H. Kitzman, R. Cole, J. Robinson, K. Sidora, D. W. Luckey, C. R. J. Henderson, C. Hanks, J. Bondy, and J. Holmberg (2004). Effects of nurse home-visiting on maternal life course and child development: Age 6 follow-up of a randomized trial. *Pediatrics 114*(6), 1550–1559.

Olds, D. L., H. Kitzman, C. Hanks, R. Cole, E. Anson, K. Sidora-Arcoleo, D. W. Luckey, C. R. J. Henderson, J. Holmberg, R. A. Tutt, A. J. Stevenson, and J. Bondy (2007). Effects of nurse home visiting on maternal and child functioning: Age-9 follow-up of a randomized trial. *Pediatrics 120*(4), e832–e845.

Olds, D. L., J. Robinson, R. O'Brien, D. W. Luckey, L. M. Pettitt, C. R. J. Henderson, R. K. Ng, K. L. Sheff, J. Korfmacher, S. Hiatt, and A. Talmi (2002). Home visiting by paraprofessionals and by nurses: A randomized controlled trial. *Pediatrics 110*(3), 486–496.

Petek, N. and N. G. Pope (2023). The multidimensional impact of teachers on students. *Journal of Political Economy 131*(4), 1057–1107.

Robling, M., M.-J. Bekkers, K. Bell, C. C. Butler, R. Cannings-John, S. Channon, B. C. Martin, J. W. Gregory, K. Hood, A. Kemp, et al. (2015). Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (building blocks): a pragmatic randomised controlled trial. *The Lancet 387*(10014), 146–155.

Robling, M., F. Lugg-Widger, R. Cannings-John, J. Sanders, L. Angel, S. Channon, D. Fitzsimmons, K. Hood, J. Kenkre, G. Moody, E. Owen-Jones, R. Pockett, J. Segrott, and T. Slater (2021). The family nurse partnership to reduce maltreatment and improve child health and development in young children: the bb:2–6 routine data-linkage follow-up to earlier rct. *Public Health Research 9*(2).

Robling, M., F. V. Lugg-Widger, R. Cannings-John, L. Angel, S. Channon, D. Fitzsimmons, K. Hood, J. Kenkre, G. Moody, E. Owen-Jones, et al. (2022). Nurse-led home-visitation programme for first-time mothers in reducing maltreatment and improving child health and development (bb: 2-6): longer-term outcomes from a randomised cohort using data linkage. *BMJ open 12*(2), e049960.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Schepan, M. L., M. Sandner, G. Conti, S. Kliem, and T. Brand (2025). Maternal and child health following 2 home visiting interventions vs control: five-year follow-up of a randomized clinical trial. *JAMA pediatrics 179*(4), 367–374.

Schonhaut, L., A. Maturana, O. Cepeda, and P. Serón (2021). Predictive validity of developmental screening questionnaires for identifying children with later cognitive or educational difficulties: A systematic review. *Frontiers in Pediatrics 24*.

Schonhaut, L., M. Perez, I. Armijo, , and A. Maturana (2020). Comparison between *a*ges & *s*tages *q*uestionnaire and *b*ayley *s*cales, to predict cognitive delay in school age. *41*(104933).

Steenland, M. W., D. Oviedo, M. A. Bates, A. Zhou, C. Zera, K. Baicker, and M. A. McConnell (2025). Effect of an intensive nurse home visiting program on postpartum contraceptive use and birth spacing: a randomized controlled trial. *Obstetrics & Gynecology 145*(1), 3–12.

Stoye, G. (2025). The distribution of doctor quality: Evidence from cardiologists in England. *Journal of Political Economy Microeconomics*.

Swanson, K., S. Ayers, D. Oviedo, R. A. Zhou, M. A. Bates, K. Baicker, A. T. Chien, and M. McConnell (2025). Intensive nurse home visiting and early childhood outcomes: a randomized clinical trial. *JAMA pediatrics 179*(8), 857–866.

Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement*. *The Economic Journal 113*(485), F3–F33.

Wüst, M. and E. Sørensen (2025). Good nurses. Technical report, Unpublished manuscript.

Zaranko, B., N. J. Sanford, E. Kelly, A. M. Rafferty, J. Bird, L. Mercuri, J. Sigsworth, M. Wells, and C. Propper (2023). Nurse staffing and inpatient mortality in the english national health service: a retrospective longitudinal study. *BMJ quality & safety 32*(5), 254–263.

Table 1: Mother's characteristics at FNP enrolment

	(1) Mean	(2) Std Dev.	(3) Min	(4) Max
Demographics				
Age (years)	17.83	1.36	12.7	24.3
Gestational age (weeks)	18.02	5.00	5	40
Primary language English	0.96	0.20	0	1
Ethnicity: Black	0.06	0.23	0	1
Ethnicity: Asian	0.03	0.16	0	1
Ethnicity: Mixed/Other	0.07	0.25	0	1
Social and family indicators				
Mother has partner	0.77	0.42	0	1
Contact with bio dad: Never	0.14	0.34	0	1
Contact with bio dad: < Weekly	0.06	0.24	0	1
Contact with bio dad: Weekly	0.10	0.31	0	1
Contact with bio dad: Daily	0.49	0.50	0	1
Lives with own partner	0.28	0.45	0	1
Lives with own mother	0.55	0.50	0	1
Economic indicators				
NEET	0.54	0.50	0	1
Lacks good qualifications	0.69	0.46	0	1
Any vocational qualification	0.41	0.49	0	1
Ever in paid work	0.44	0.50	0	1
Current in paid work	0.16	0.37	0	1
Income only from benefits	0.32	0.47	0	1
Residence: Private renter	0.21	0.40	0	1
Residence: Social renter	0.46	0.50	0	1
Residence: Non-response	0.15	0.36	0	1
Residence: Non-traditional	0.06	0.24	0	1
Social services				
Any social care	0.14	0.35	0	1
Any social services	0.51	0.50	0	1

Note: Summary statistics of characteristics of mothers ever enrolled in the FNP between 2007 and 2013. Sample size varies slightly across variables, between 15,069 and 16,436. NEET: Not in Employment, Education or Training.

Table 2: Family nurse characteristics

	(1) Mean	(2) Std. Dev	(3) Min	(4) Max
Demographics				
Female	0.99	0.12	0	1
Non-white ethnicity	0.07	0.26	0	1
Age: 40 or older	0.76	0.43	0	1
Age: 50 or older	0.28	0.45	0	1
Qualifications				
Nurse - general	0.56	0.50	0	1
Nurse - adult	0.22	0.42	0	1
Nurse - mental health	0.05	0.22	0	1
Nurse - children	0.21	0.41	0	1
Nurse - learning disabilities	0.01	0.08	0	1
Midwife	0.39	0.49	0	1
School nurse	0.03	0.18	0	1
Certification in health visiting	0.17	0.38	0	1
BSc in school nursing or equivalent	0.09	0.29	0	1
BSc in health visiting or equivalent	0.56	0.50	0	1
MSc credits	0.23	0.42	0	1
MSc degree	0.10	0.30	0	1
FN qualification: adult and/or general	0.76	0.43	0	1
Previous experience				
Working in the community	0.97	0.16	0	1
Working with teenage parents	0.74	0.44	0	1
Named Nurse Safeguarding	0.03	0.16	0	1
Child/adolescent mental health services	0.11	0.31	0	1
Additional training				
Working together safeguarding training	0.52	0.50	0	1
Safeguarding supervision training	0.18	0.38	0	1
Common Assessment Framework training	0.66	0.48	0	1
Management training	0.11	0.31	0	1
Other additional training	0.24	0.43	0	1
Role in the team				
Supervisor	0.17	0.38	0	1

Note: Summary statistics of characteristics of family nurses ever observed working in the FNP between 2007 and 2013. Sample size varies slightly across variables, between 326 and 472.

Table 3: Tests of within-site quasi-random assignment of family nurses to clients, sample of mothers enrolled in the FNP between 2007 and 2013

	# outcomes	# outcomes where mother Xs significant		
		No MHT	HBS	RW
A - Benchmark specification				
All FN characteristics	38	9	0	0
% of all FN characteristics		23.7%	0.0%	0.0%
FN demographics	5	0	0	0
FN qualifications	12	0	0	0
FN experience	5	1	0	0
FN training	5	2	0	0
Process quality	11	6	0	0
B - Not controlling for case complexity				
All FN characteristics	38	9	0	0
% of all FN characteristics		23.7%	0.0%	0.0%
FN demographics	5	0	0	0
FN qualifications	12	0	0	0
FN experience	5	1	0	0
FN training	5	2	0	0
Process quality	11	6	0	0
C - Not controlling for case complexity or mother referral variables				
All FN characteristics	38	11	0	0
% of all FN characteristics		28.9%	0.0%	0.0%
FN demographics	5	0	0	0
FN qualifications	12	0	0	0
FN experience	5	1	0	0
FN training	5	1	0	0
Process quality	11	9	0	0

Note: The table reports the results of the test of quasi-random assignment of clients to FN on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the FN characteristics in a regression of a FN characteristic on mother characteristics not known at the time of assignment, controlling for those known (mother's age, ethnicity, whether she speaks English, baby's gestational age at enrollment), the FN caseload size and complexity at the time of assignment, site fixed effects, and mother-year enrollment dummies. Standard errors are clustered at the site level. The table reports the number of FN characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test is rejected, when the p-value is not adjusted for multiple hypothesis testing in column (3), where it is adjusted using the Holm-Bonferroni-Sidak adjustment in column (4), and where it is adjusted using the Romano-Wolf procedure in column (5). Column (1) reports the total number of FN characteristics. In each panel, the first two rows refer to the entire set of 38 FN characteristics we test for, while the rows below report the results by groups of FN characteristics.

Table 4: Distribution of family nurse effectiveness for different child outcomes

	Child outcome				
	Birth outcomes	Cognitive age 1	Cognitive age 2	Socio-emo. age 1	Socio-emo. age 2
Panel A - Within sites					
SD	0.109	0.218	0.196	0.307	0.234
p90 - p10	0.288	0.547	0.488	0.761	0.588
p75 - p25	0.141	0.279	0.232	0.37	0.307
Panel B - Across sites					
SD	0.121	0.266	0.239	0.371	0.292
p90 - p10	0.311	0.682	0.628	0.908	0.75
p75 - p25	0.154	0.363	0.298	0.46	0.4
No. of FNs	647	616	585	611	574
No. of mothers	18,258	13,820	10,383	13,060	9,700

Note: The table reports the standard deviation of the distribution of family nurse fixed effects, estimated as in Equation 3. The sample includes mothers who joined the FNP by December 2013 and have a valid outcome measure, and FNs who have at least five mothers. In Panel A, outcomes are first demeaned within site to account for fixed differences across FNP sites. We apply a Bayesian shrinkage estimator to all results. Source: Authors' calculations using data from the Department of Health and Social Care.

Table 5: Robustness of estimates of distribution of family nurse effectiveness to different controls

	Birth outcomes	Cognitive Age 1	Socio-emot. Age 1	Cognitive Age 2	Socio-emot. Age 2
A - Benchmark (Within sites)					
Standard deviation	0.109	0.218	0.307	0.196	0.234
Number of FNs	647	616	611	585	574
Number of mothers	18,258	13,820	13,060	10,383	9,700
B - Dropping controls for caseload complexity					
Standard deviation	0.109	0.218	0.306	0.196	0.233
Number of FNs	647	616	611	585	574
Number of mothers	18,258	13,820	13,060	10,383	9,700
C - Dropping controls for caseload complexity & mother demographics					
Standard deviation	0.112	0.217	0.305	0.194	0.233
Number of FNs	647	616	611	585	574
Number of mothers	18,258	13,820	13,060	10,383	9,700
D - Controlling for extended set of mother characteristics collected at enrollment					
Standard deviation	0.108	0.217	0.306	0.198	0.231
Number of FNs	647	616	611	585	574
Number of mothers	18,258	13,820	13,060	10,383	9,700
E - Common sample across items					
Standard deviation	0.142	0.228	0.302	0.194	0.237
Number of FNs	559	559	559	559	559
Number of mothers	9,098	9,098	9,098	9,098	9,098
F - At least three mothers per family nurse (item-specific sample)					
Standard deviation	0.111	0.221	0.316	0.215	0.243
Number of FNs	658	658	651	635	624
Number of mothers	18,298	13,964	13,201	10,556	9,866
G - At least ten mothers per family nurse (item-specific sample)					
Standard deviation	0.099	0.215	0.287	0.190	0.221
Number of FNs	586	542	529	472	442
Number of mothers	17,827	13,300	12,486	9,583	8,760
H - Excluding sites with sustained supervisor overlap					
Standard deviation	0.114	0.221	0.319	0.201	0.240
Number of FNs	501	473	470	448	435
Number of mothers	13,748	10,402	9,820	7,752	7,181

Note: The table reports the standard deviation of the distribution of family nurse fixed effects, estimated as in Equation 3. Outcomes are first demeaned within site to account for fixed differences across FNP sites. The sample includes mothers who joined the FNP by December 2013 and have a valid outcome measure, and FNs who have at least five clients. In Panels B and C, we omit vectors of control variables related to caseload size and complexity and basic mother demographics (age at enrolment, gestational age at enrolment, whether mother's primary language is English, and maternal ethnicity). In Panel D, we reinstate these controls and additionally control for the extended set of mother characteristics collected as part of the enrolment process (listed in Table 1). In Panel E, we impose a common sample across outcomes. In Panels F and G, we return to item-specific samples but vary the minimum number of valid clients that family nurses must have to be included in the estimation. In Panel H, we exclude the 24 sites (18.6% of the sample) in which two or more supervisors were simultaneously active for two or more months, as identified using the supervisor tenure panel described in footnote 4. We apply a Bayesian shrinkage estimator to all results.

Table 6: Correlations between family nurse effectiveness at promoting different child outcomes

	Birth	Cognitive Age 1	Cognitive Age 2	Socio-emotional Age 1	Socio-emotional Age 2
	(1)	(2)	(3)	(4)	(5)
(1) Birth outcomes		**0.165**	**0.143**	0.004	0.067
(2) Cognitive, age 1	**0.165**		**0.671**	**0.338**	**0.338**
(3) Cognitive, age 2	**0.143**	**0.671**		**0.254**	**0.373**
(4) Socio-emot., age 1	0.004	**0.338**	**0.254**		**0.575**
(5) Socio-emot., age 2	0.067	**0.338**	**0.373**	**0.575**	

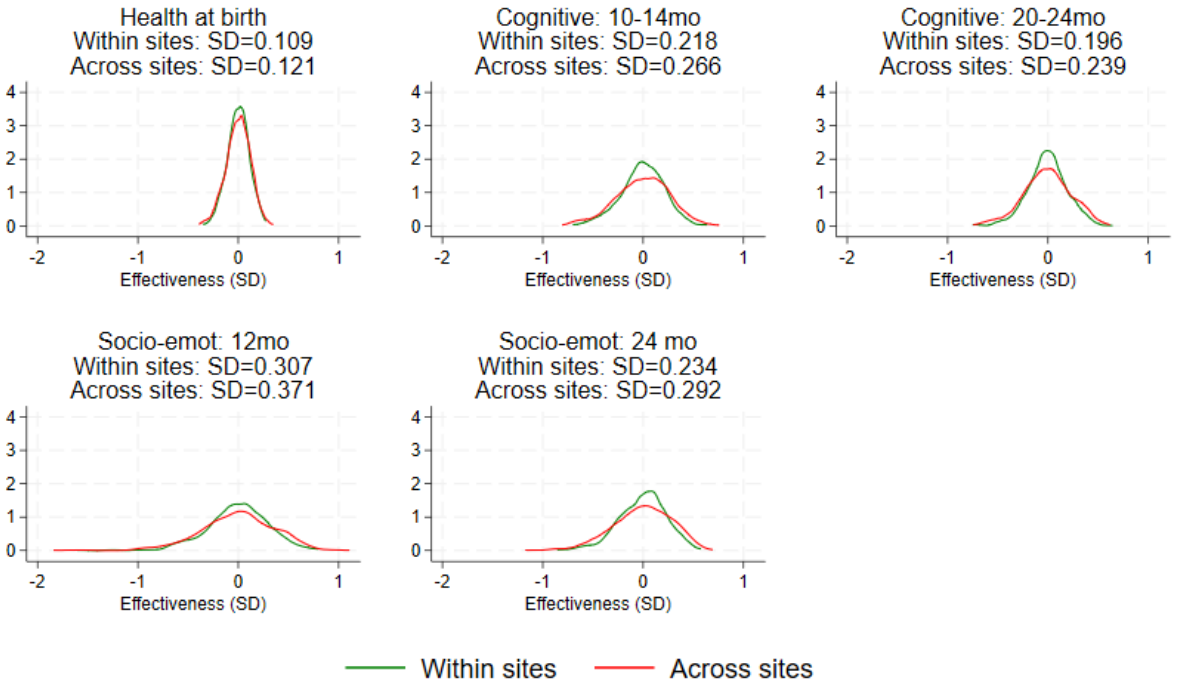
Note: Correlations in **bold** are statistically significant at the 5% level.

Table 7: Impact of family nurse effectiveness on maternal behaviours and investments in children

	Dependent variable:				
	Maternal risky behaviours	Maternal socio-economic circum-stances	Maternal relationships	Maternal mental health	Investments in child development
FN effectiveness at cognitive development, age 1	-0.130	0.085	0.083	0.075	0.008
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
p-val	*0.002*	*0.043*	*0.049*	*0.074*	*0.853*
Adjusted p-val	*0.010*	*0.184*	*0.184*	*0.184*	*0.864*
N	11,427	11,427	11,427	11,427	11,427
FN effectiveness at cognitive development, age 2	-0.117	-0.009	0.057	-0.074	0.123
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
p-val	*0.020*	*0.859*	*0.257*	*0.140*	*0.015*
Adjusted p-val	*0.076*	*0.840*	*0.453*	*0.371*	*0.076*
N	9,798	9,798	9,798	9,798	9,798
FN effectiveness at socio-emot. development, age 1	-0.039	0.038	0.033	0.091	0.059
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
p-val	*0.218*	*0.229*	*0.291*	*0.004*	*0.060*
Adjusted p-val	*0.521*	*0.521*	*0.521*	*0.020*	*0.210*
N	11,274	11,274	11,274	11,274	11,274
FN effectiveness at socio-emot. development, age 2	-0.041	0.103	-0.012	0.104	0.005
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
p-val	*0.370*	*0.025*	*0.802*	*0.024*	*0.906*
Adjusted p-val	*0.762*	*0.126*	*0.964*	*0.126*	*0.964*
N	9,169	9,169	9,169	9,169	9,169

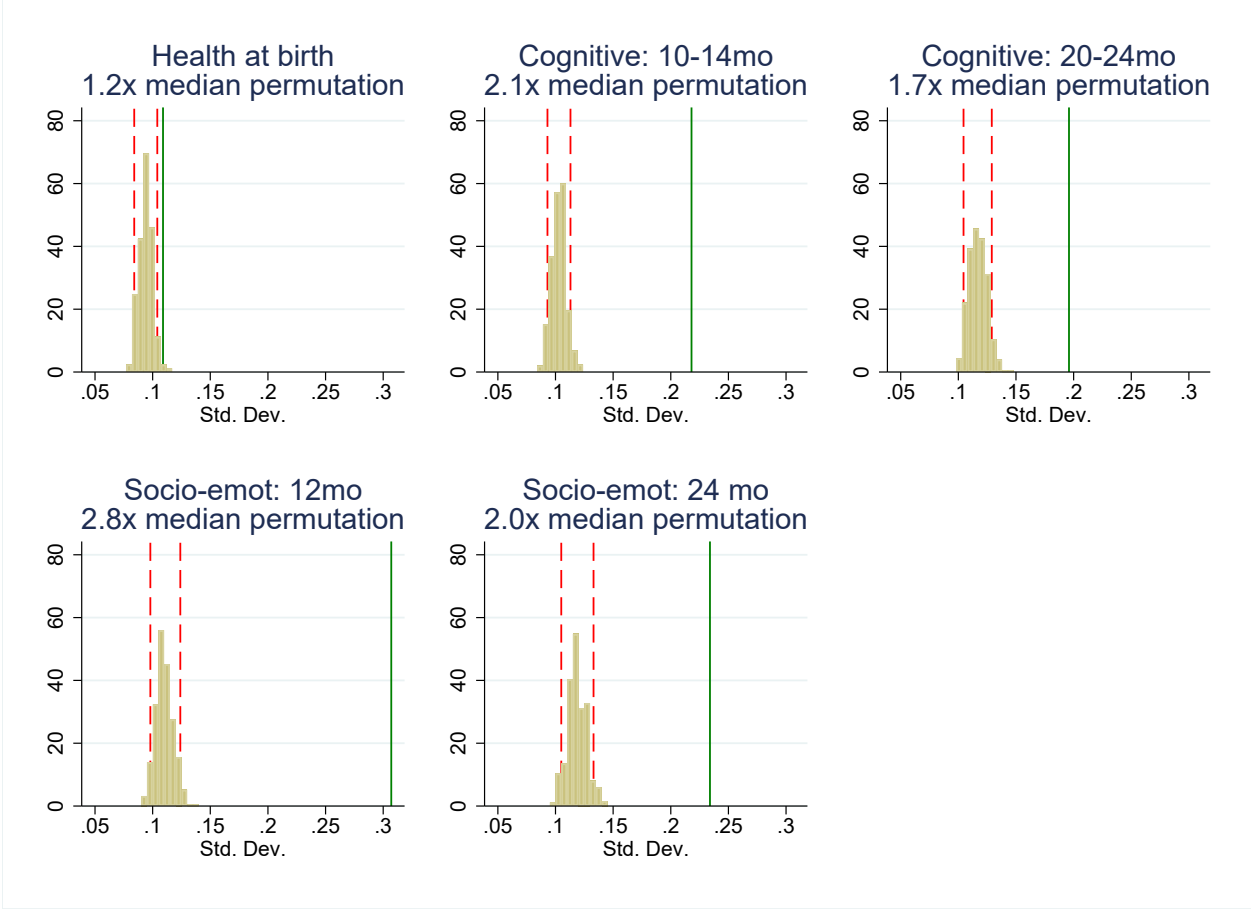
Note: Each coefficient comes from a linear regression of the mechanism on the relevant measure of family nurse effectiveness (calculated using the leave-out method). Under each coefficient we report its standard error, p-value and p-value adjusted for multiple hypotheses testing across the five outcomes following Romano and Wolf (2005). Mechanisms are standardised to have mean zero and standard deviation one, so coefficient magnitudes are comparable across specifications. Mechanisms are constructed using relevant measures up to the relevant age; for example, 'age 1' risky behaviours includes information on smoking, drinking, and marijuana use at 36 weeks of pregnancy; new sexually transmitted infections and urinary tract infections between FNP enrolment and the child's birth; and smoking, drinking, marijuana use, and unprotected sexual intercourse when the child was 12 months old. The 'age 2' risky behaviour factor adds unprotected sexual intercourse in the six months leading up to the child's second birthday (the end of the intervention).

Figure 1: Distribution of family nurse effectiveness



Note: The figure shows the distribution of estimated family nurse effectiveness, both within sites (our preferred specification, as estimated in Equation 3) and across sites (Equation 6). Each outcome is standardised to have mean 0 and standard deviation 1. The sample includes mothers who joined the FNP by December 2013 and have a valid outcome measure, and FNs who have at least five mothers included. We apply a Bayesian shrinkage estimator to all results. Source: Authors' calculations using data from the Department of Health and Social Care.

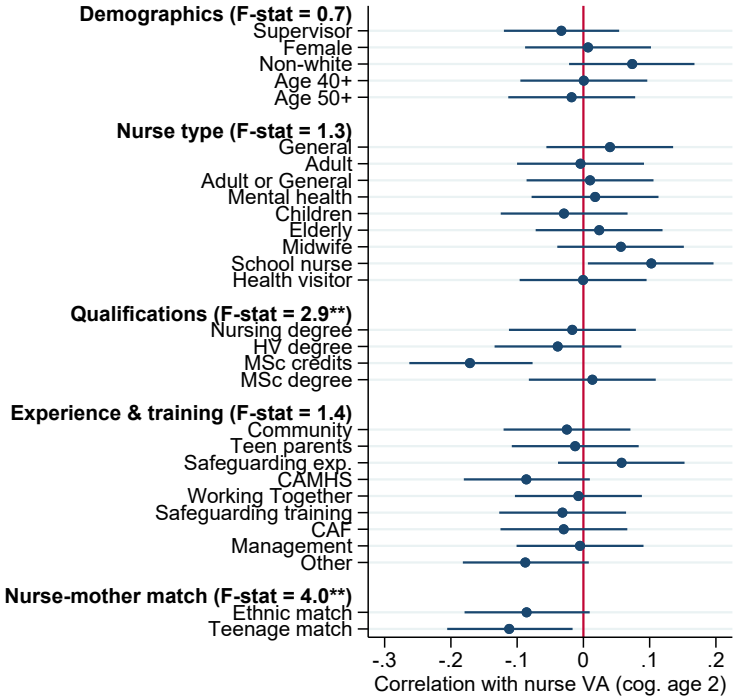
Figure 2: Randomisation inference test of the degree of heterogeneity in family nurse effectiveness



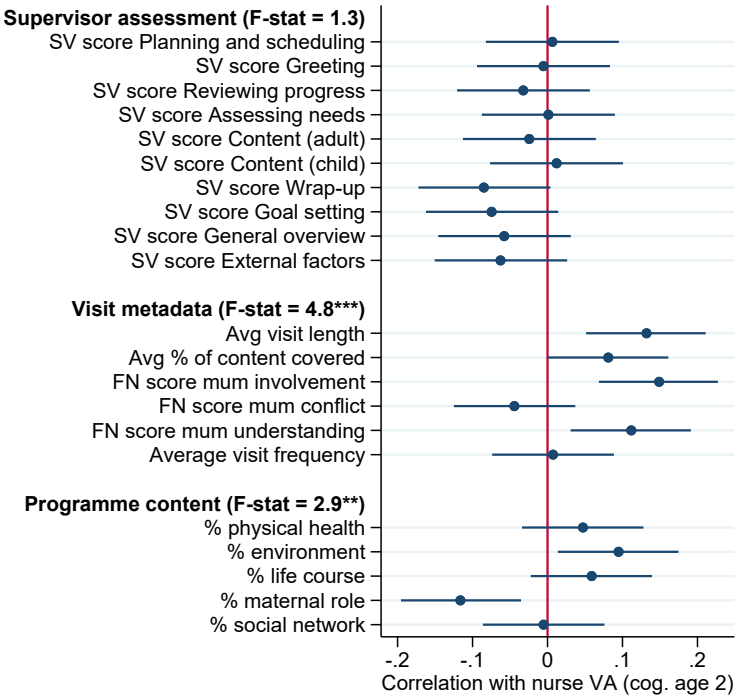
Note: In beige, we plot the distribution of the standard deviation of family nurse fixed effects, obtained estimating Equation 3 after randomly reassigning FNs to families within sites, through 500 permutations of this procedure per outcome. Red dashed vertical lines indicate the 5th and 95th percentiles of this randomisation inference distribution. Green vertical lines show the true standard deviation of the distribution of family nurse effectiveness (as in Table D.8). At the top of each figure, we compare the magnitude of the 'true' standard deviation against the median of our permutation distribution. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure 3: Pairwise correlations between family nurse effectiveness at promoting age 2 cognitive development and FN quality indicators

(a) Family Nurse 'hiring' characteristics (structural quality indicators)



(b) Family Nurse process quality measures



Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlations between family nurse effectiveness and each FN quality indicator, as specified. The F-statistic is based on a regression of family nurse effectiveness on groups of characteristics, as specified in the figure. Due to missingness in FN characteristics, these results are based on 419 nurses. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Workforce Quality and Early Childhood Development at Scale

Sarah Cattan, Gabriella Conti and Christine Farquharson

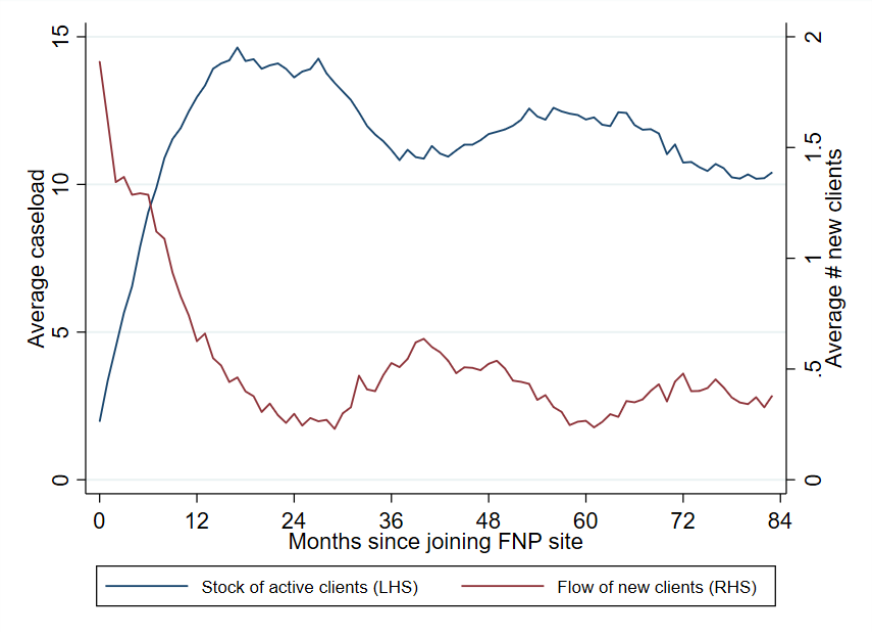
April 2026

ONLINE APPENDIX

Acknowledgements: The authors gratefully acknowledge funding from the Tavistock and Portman NHS Foundation Trust, the ESRC Centre for the Microeconomic analysis of Public Policies at the Institute for Fiscal Studies, and the NORFACE DIAL for the project "Growing up Unequal? The Origins, Dynamics and Lifecycle Consequences of Childhood Inequalities". Conti acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No.819752 DEVORHBIOSHIP – ERC-2018-COG) and from The Leverhulme Trust, via a Philip Leverhulme Prize. We are indebted to Ruth Rothman, Phil Howlin and Andreea Moise from the Family Nurse Partnership National Unit for answering our numerous questions about the data and the programme. We thank seminar participants at the "Medical Interventions and Socioeconomic Outcomes" Workshop at Odense University, UCL Research Day Workshop, EALE, EEA, ESPE conferences, the Economic Demography Workshop, the Paris School of Economics workshop on early childhood development, the Workshop on Labour and Family Economics at the University of York, the NORFACE-DIAL conference, the Barcelona Summer Forum, IBEO Alghero, NBER Summer Institute Children Group, Bank of Italy Human Capital Workshop, Workforce Workshop at the University of Surrey, Opportunity & Inclusive Growth Institute, and at the University of Copenhagen, Venice Ca' Foscari, Queen Mary, Bologna, Padova, Princeton, Naples, Oxford, Warwick, Aarhus, LSE, Zurich, Bocconi, PUC, University of Chile, UCLA, UCSB, NYUAD, York, Bergen, Berlin, CESR USC, UC Davis, Rotterdam, Copenhagen Business School, and CESifo for useful comments. All errors remain our own. Cattan: IFS [sarah_c@ifs.org.uk]; Conti: UCL [gabriella.conti@ucl.ac.uk]; Farquharson: IFS [christine_f@ifs.org.uk].

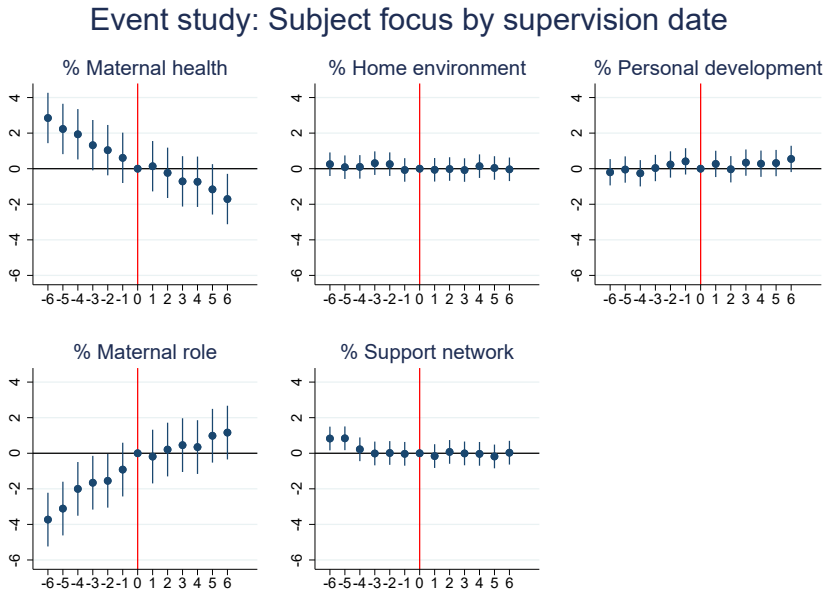
A Appendix Figures and Tables

Figure A.1: Family nurse caseloads by tenure in the FNP



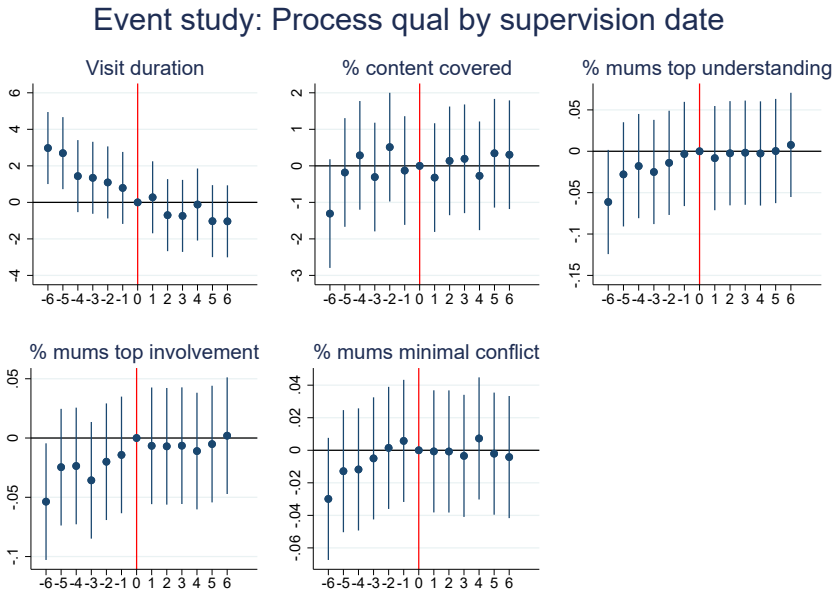
Note: The figure shows the average caseload (number of active clients) and the average number of new clients assigned to a family nurse, by the number of months since she joined the site. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.2: Visit content domain by time to/since a supervision visit



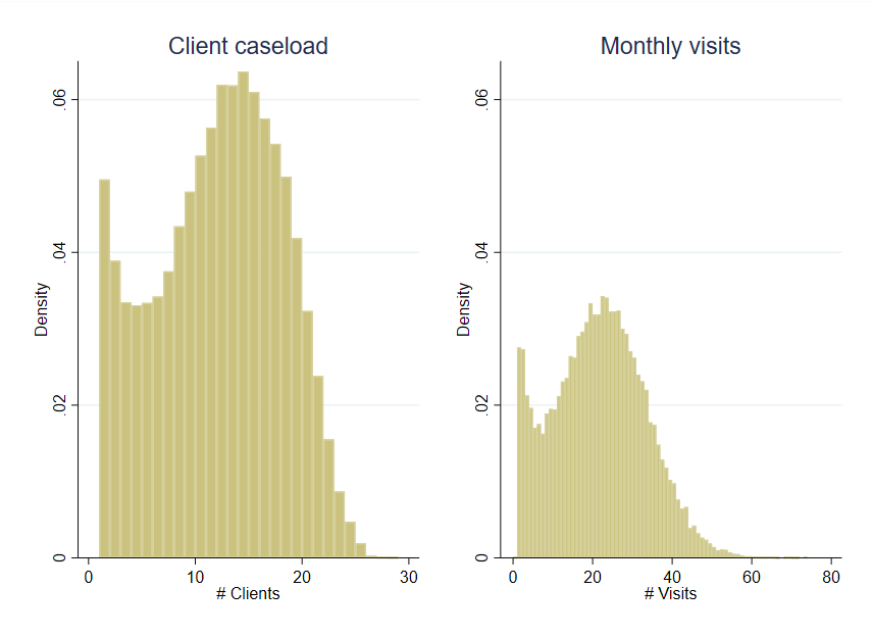
Note: The figure shows the coefficients from event study regressions of visit content domains on the months to/since the family nurse has a supervisory visit. We exclude cases where the FN has another supervisory visit within 6 months. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.3: Process quality indicators by time to/since a supervision visit



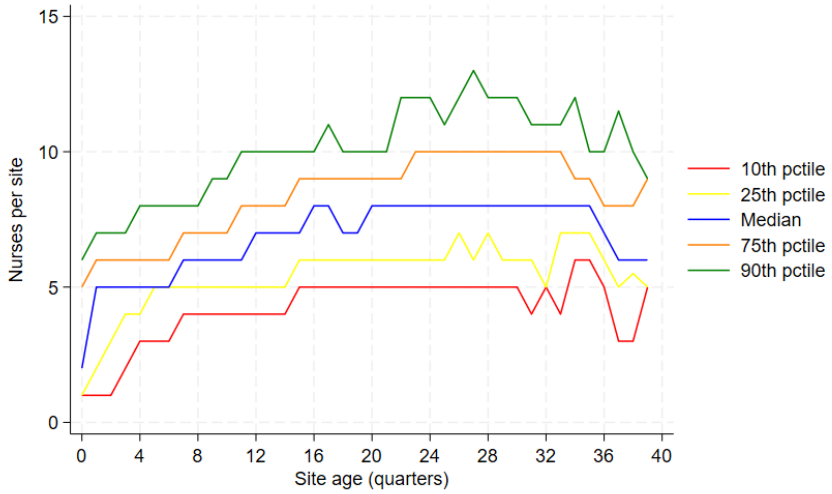
Note: The figure shows the coefficients from event study regressions of process quality indicators on the months to/since the FN has a supervisory visit. We exclude cases where the FN has another supervisory visit within 6 months. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.4: Monthly distribution of visits



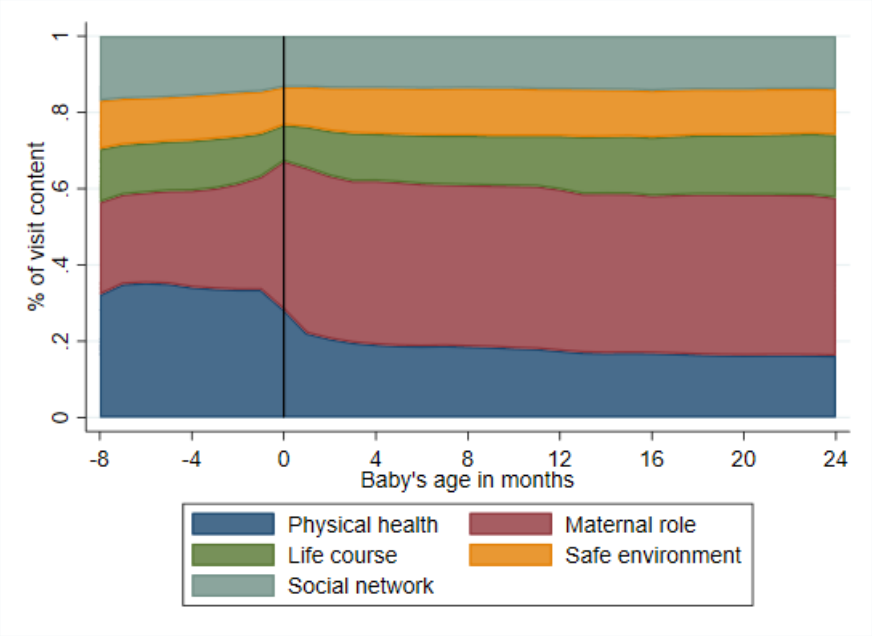
Note: The left panel shows the distribution of number of clients per family nurse in any month (based on a panel dataset at the month-family nurse level, for month-family nurse pairs where a family nurse is active in the programme). The right panel shows the distribution of number of monthly visits carried out. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.5: Number of active family nurses per site, by age of site



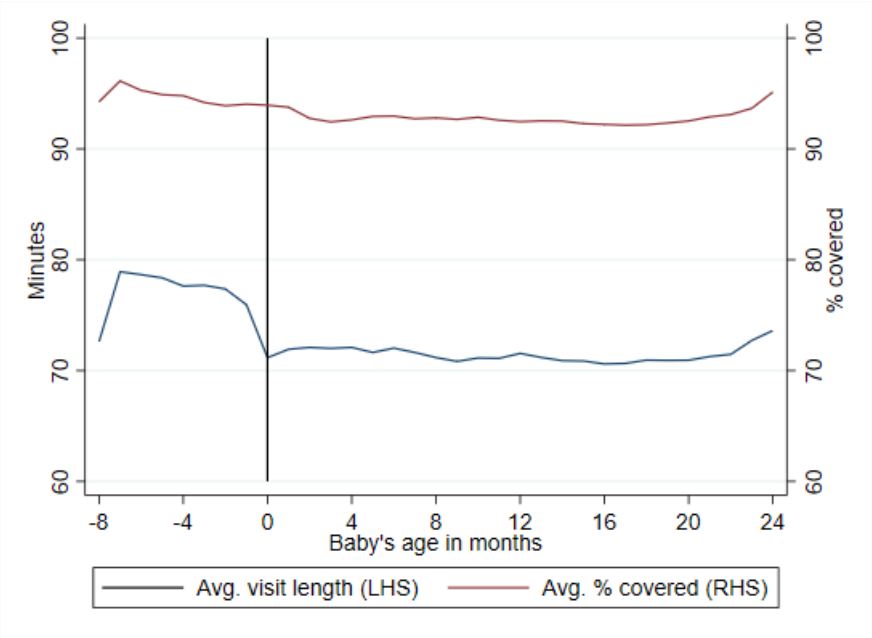
Note: We calculate each site's age based on the number of quarters since the first home visit delivered in it. For each age, we calculate the distribution of the number of family nurses active in that site. The figure presents moments from this distribution. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.6: FNP programme content, by age of child



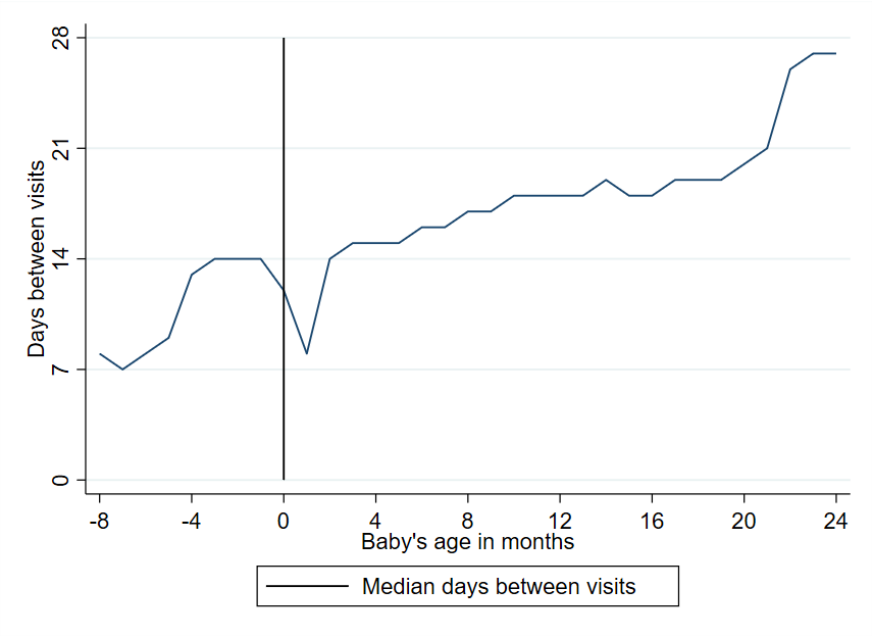
Note: The figure shows the proportion of time each visit allocated to each of the five strands of programme content, averaged across all visits. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.7: Visit duration and coverage of planned content, by age of child



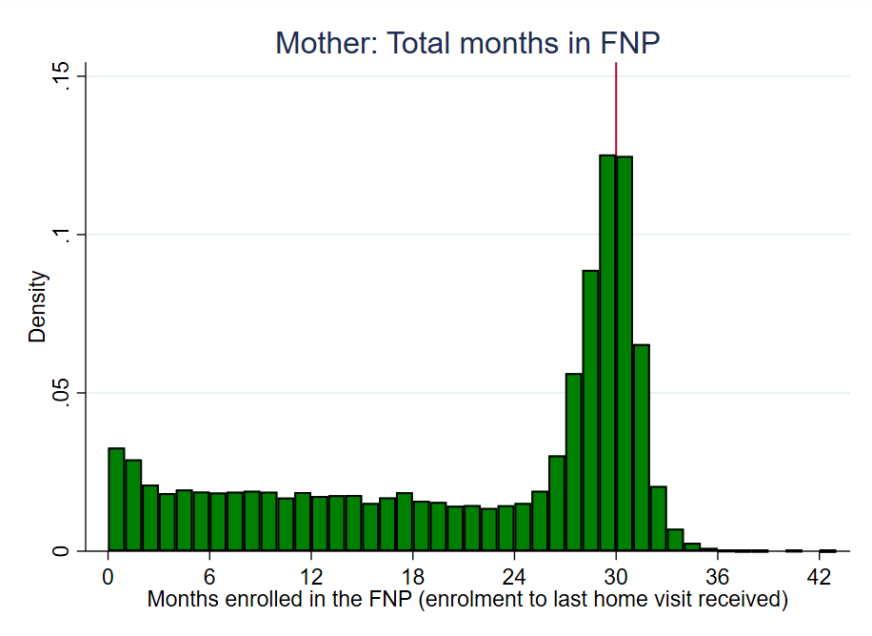
Note: The figure shows the average length of a visit (in minutes) and the average proportion of the planned content covered, by the age of the client's child. Both outcomes are recorded by the family nurse for each visit she conducts. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.8: Time (in days) in between visits by the child's age



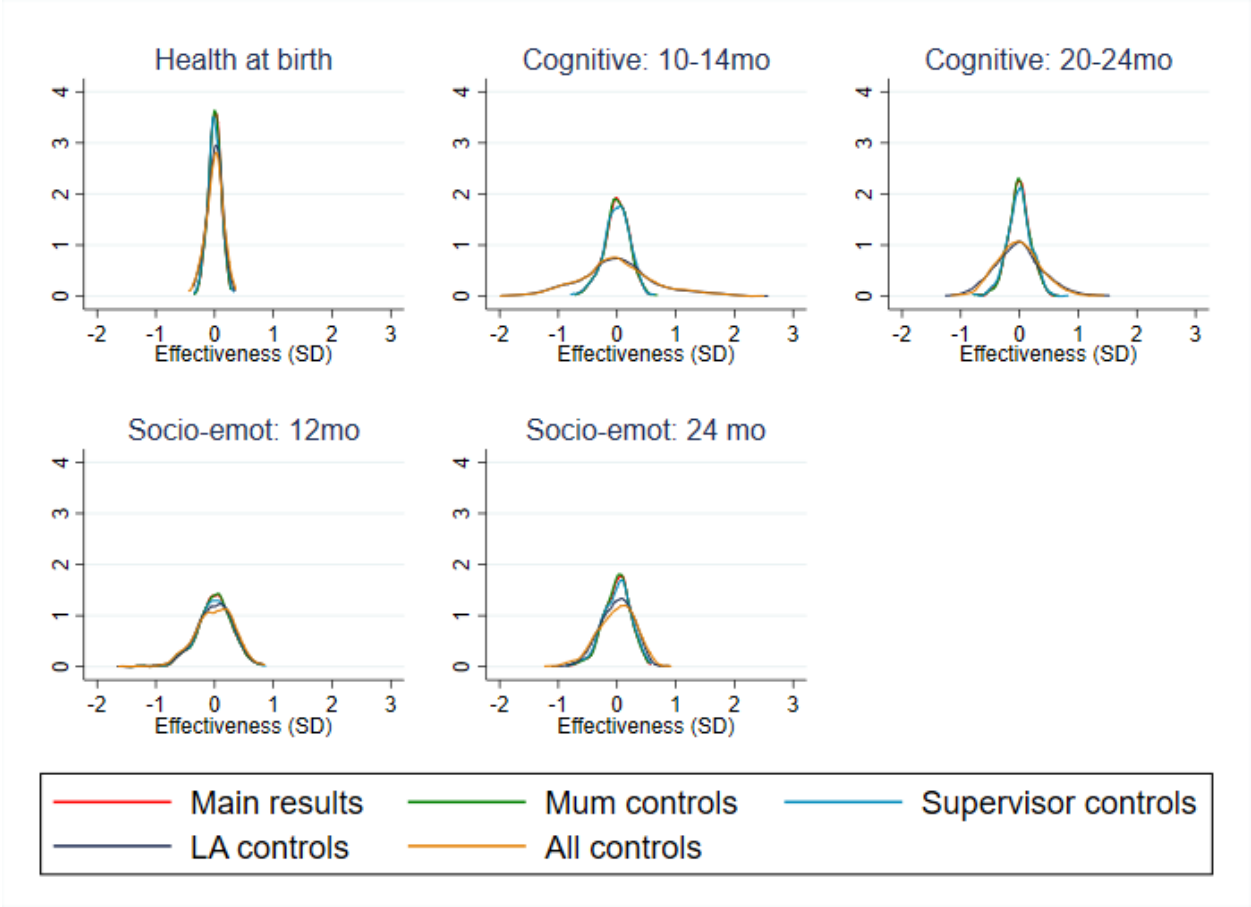
Note: Results are based on mothers who enrolled before 2014 (our analysis sample restriction). Results are based on all mothers present at each point, leading to an unbalanced panel. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.9: Distribution of mothers' total months in the FNP programme



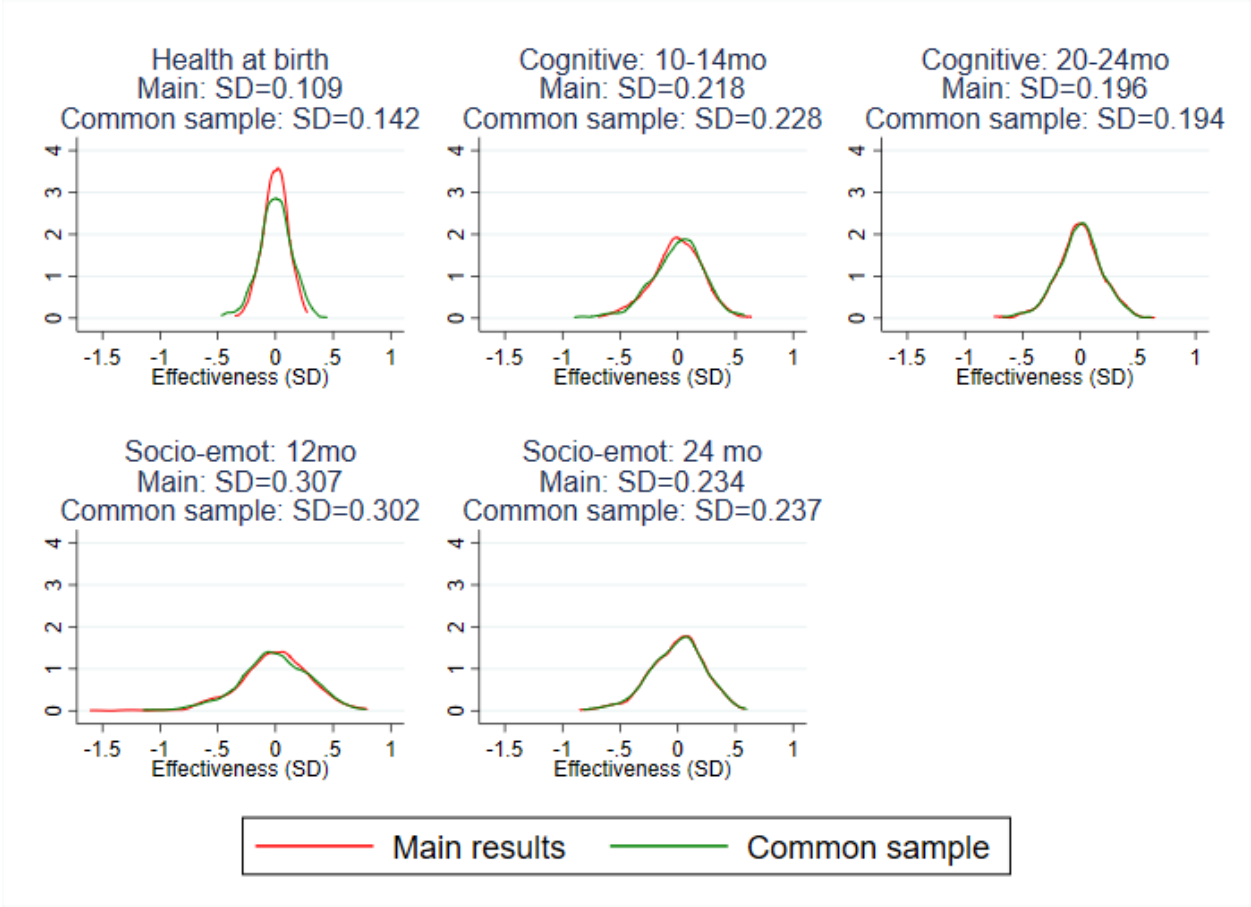
Note: On average, completion of the FNP programme requires about 30 months (six months pre-natally, and then two years after birth). Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.10: Distribution of family nurse effectiveness: Robustness



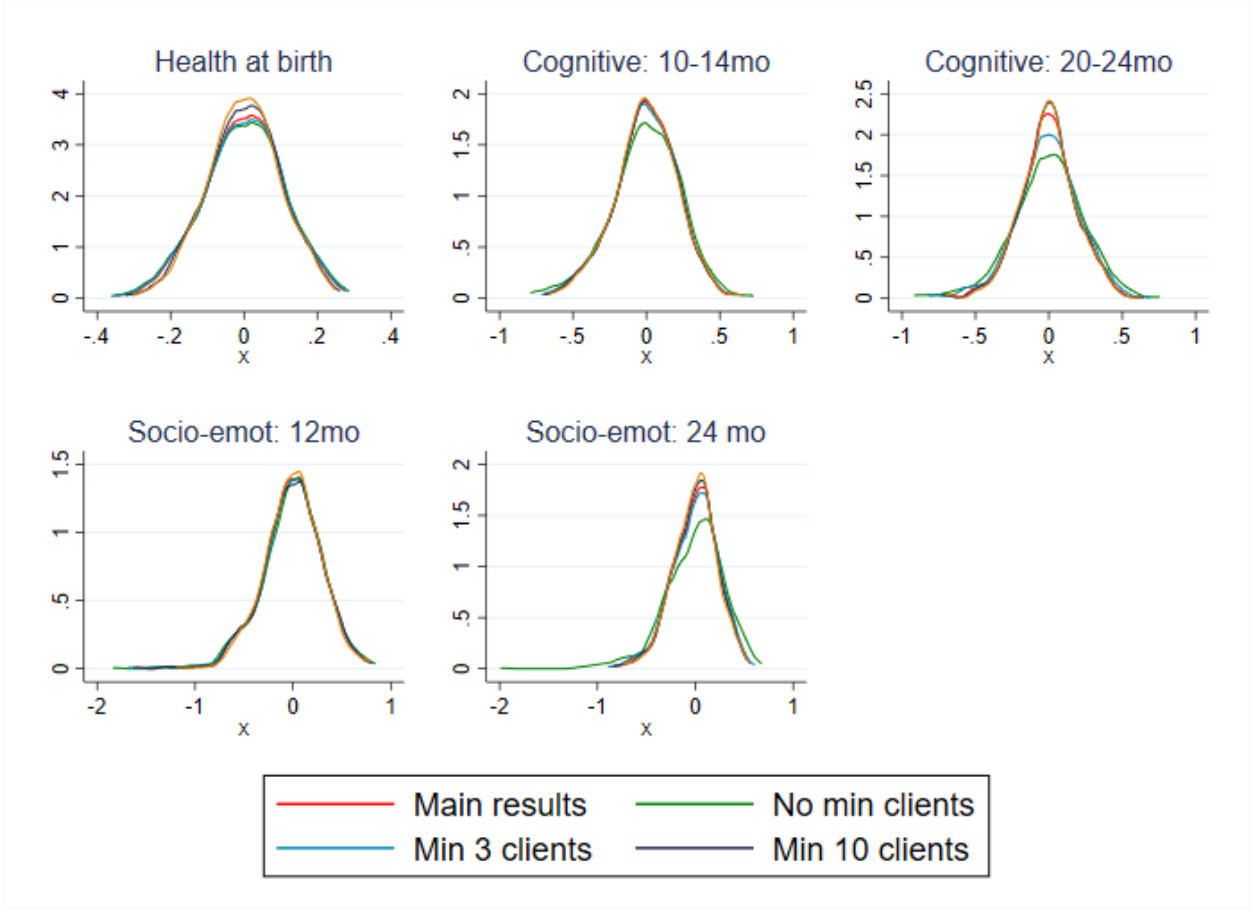
Note: See Table 5 for a description of the different control sets used.

Figure A.11: Distribution of family nurse effectiveness: Robustness to common sample



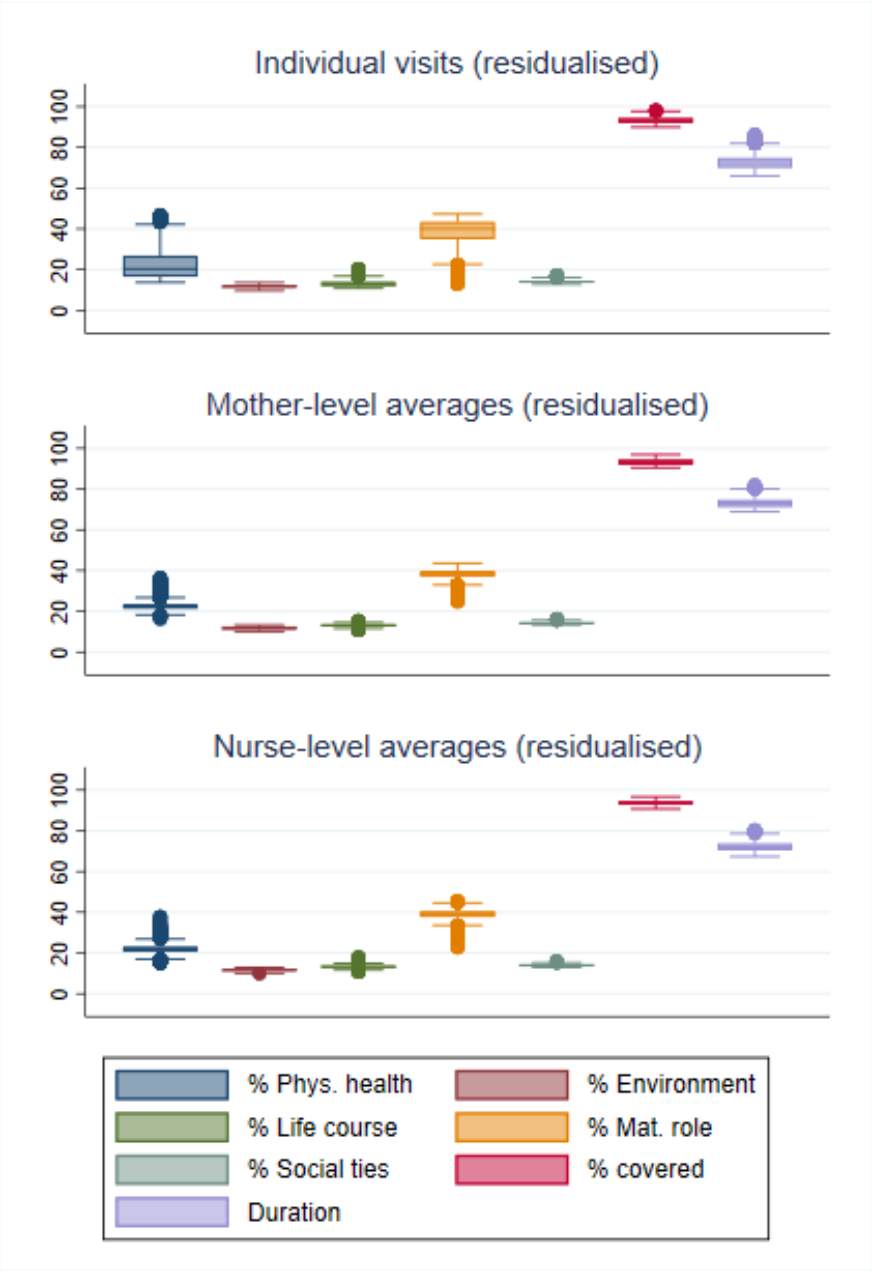
Note: The figure shows the distribution of family nurse effectiveness in promoting programme retention and each of five child outcomes. Main results are based on a sample of nurses who had joined the FNP by December 2013 and who had at least five clients with valid outcome measures. Common sample results impose a common sample across outcomes. Sample sizes are reported in Table 4. Results have been adjusted using an empirical Bayes estimator ("shrinkage estimator") to account for differences in number of clients per nurse. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.12: Distribution of family nurse effectiveness: Sensitivity



Note: See Table 5 for a description of the different control sets used.

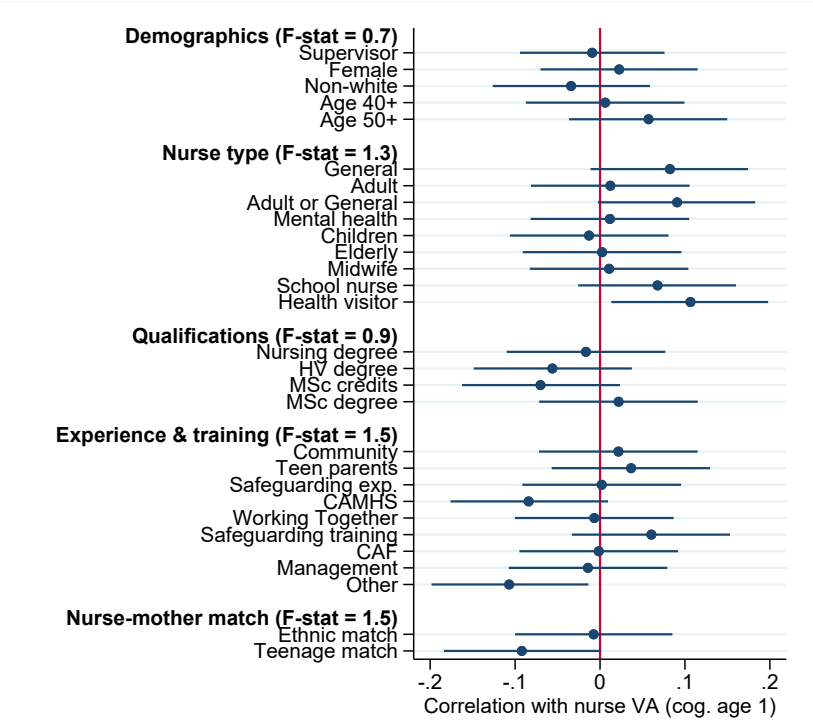
Figure A.13: Variation in visit content and duration



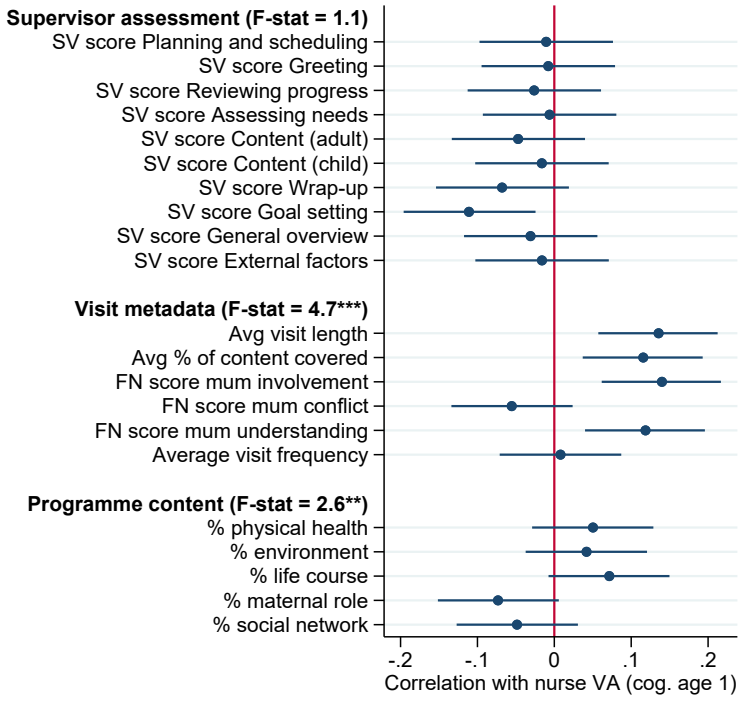
Note: The figure presents box plots for the proportion of time spent on each of five programme areas; the share of planned content covered; and visit duration. Since each dimension is related to the age of the child, we first residualise it on the child's age in months. We present these plots at three levels, showing the variation across all individual visits; the variation in the average visit received by each mother; and the variation in the average visit provided by each nurse. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.14: Pairwise correlations between family nurse effectiveness in promoting age 1 cognitive development and FN characteristics

(a) Family nurse 'hiring' characteristics



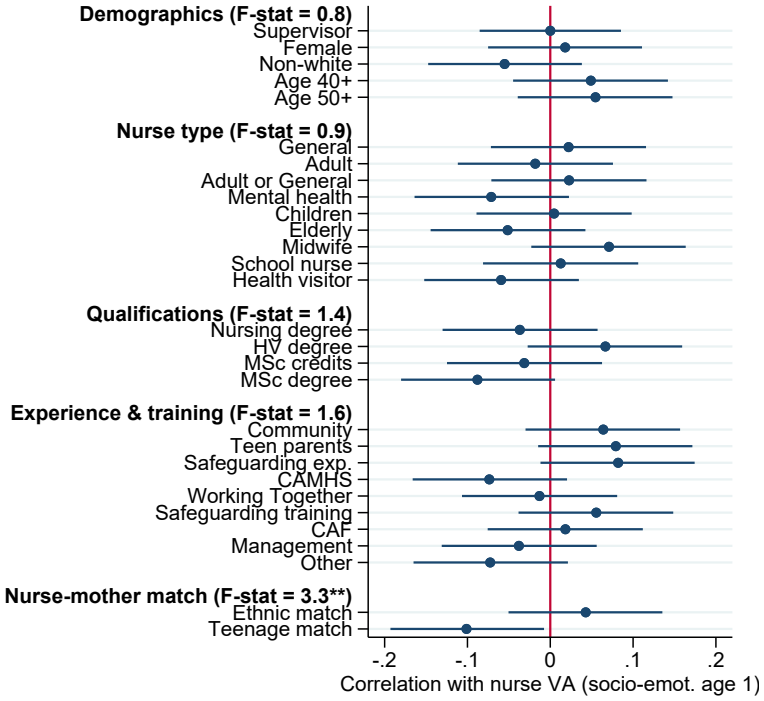
(b) Family nurse process quality measures



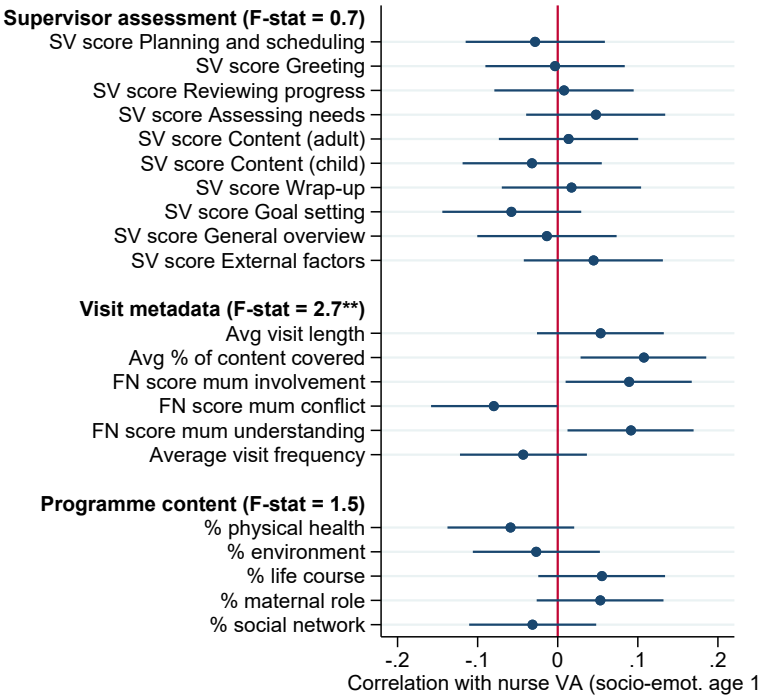
Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlation between family nurse effectiveness and each FN characteristic. The F-statistic comes from a regression of family nurse effectiveness on all characteristics in the figure. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.15: Pairwise correlations between family nurse effectiveness in promoting age 1 socio-emotional development and FN characteristics

(a) Family nurse 'hiring' characteristics



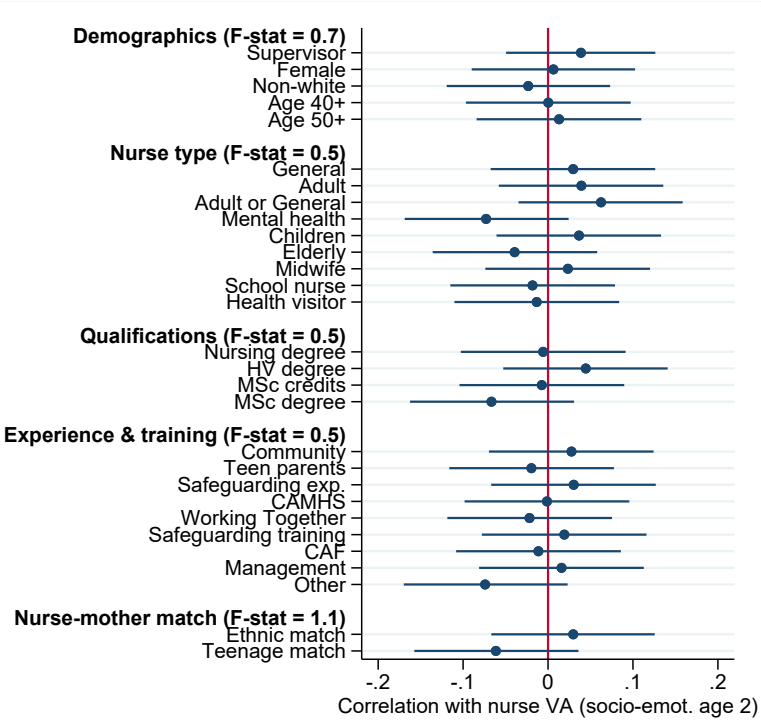
(b) Family nurse process quality measures



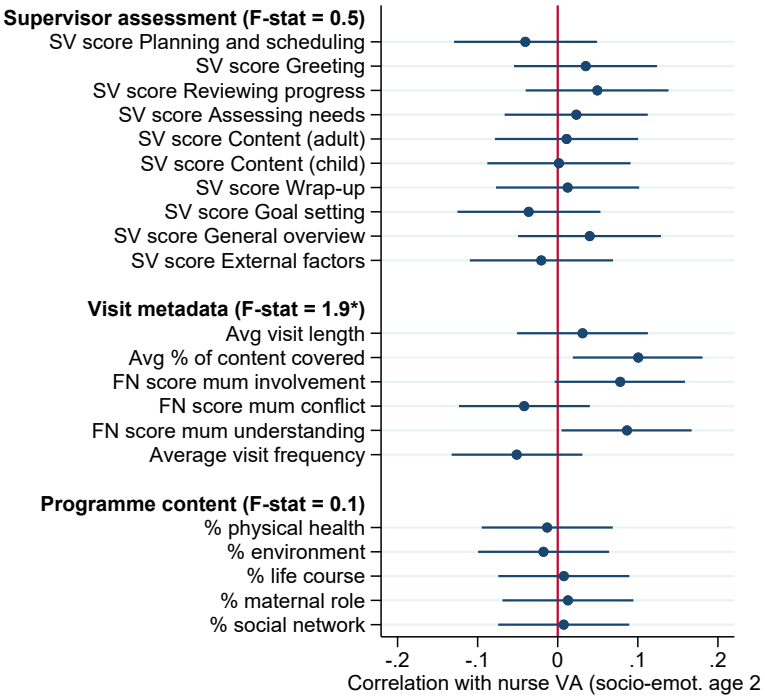
Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlation between family nurse effectiveness and each FN characteristic. The F-statistic comes from a regression of family nurse effectiveness on all characteristics in the figure. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.16: Pairwise correlations between family nurse effectiveness in promoting age 2 socio-emotional development and FN characteristics

(a) Family nurse 'hiring' characteristics

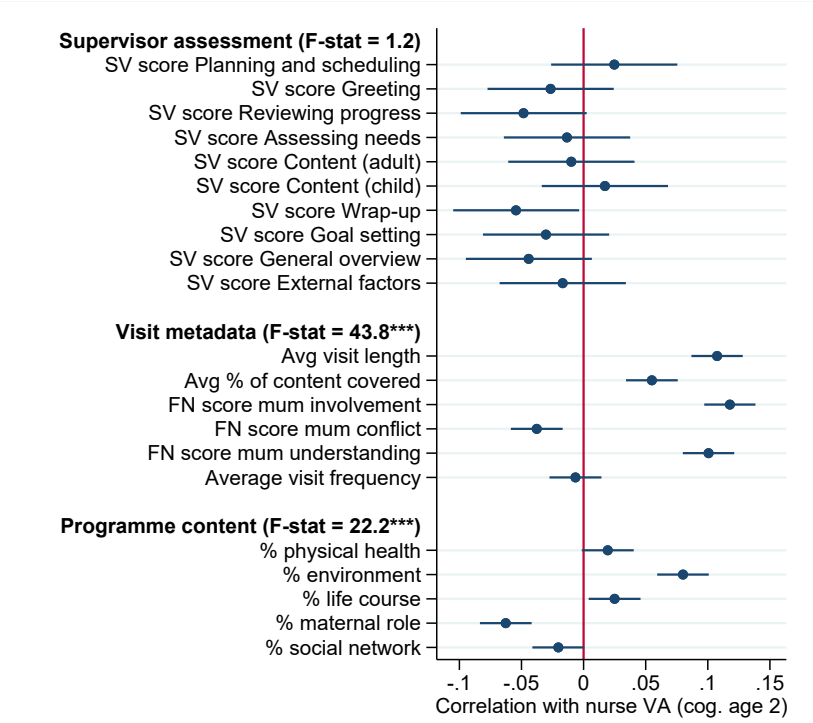


(b) Family nurse process quality measures



Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlation between family nurse effectiveness and each FN characteristic. The F-statistic comes from a regression of family nurse effectiveness on all characteristics in the figure. Source: Authors' calculations using data provided by the Department for Health and Social Care.

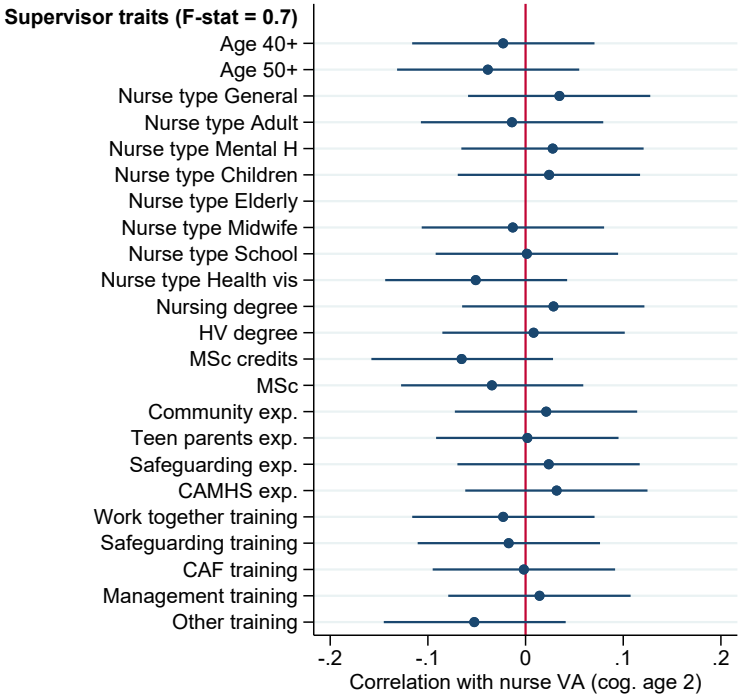
Figure A.17: Client-level pairwise correlations between leave-out measure of family nurse effectiveness at promoting age 2 cognitive development and FN process quality indicators



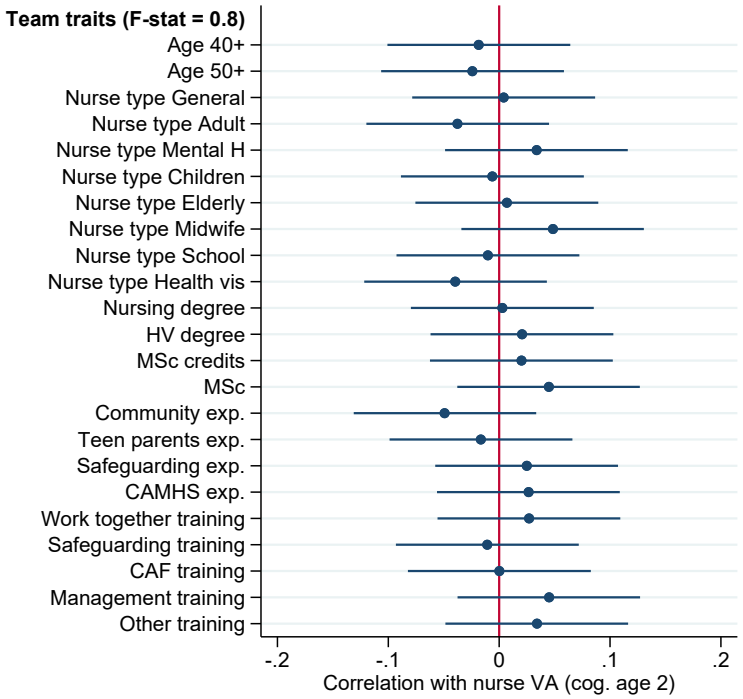
Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlation between family nurse effectiveness and each process quality characteristic. These results are based on a client-level dataset and compare the estimated family nurse effectiveness for each client (based on the leave-out method described in Jacob et al. (2010)) and the process quality measures derived from that client's records. The F-statistic comes from a regression of family nurse effectiveness on all characteristics in the figure. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure A.18: Pairwise correlations between family nurse effectiveness at promoting age 2 cognitive development and characteristics of the FN's supervisor and team

(a) Characteristics of FN's site supervisor



(b) Characteristics of FN's wider team



Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlations between family nurse effectiveness and each FN quality indicator, as specified. The F-statistic is based on a regression of family nurse effectiveness on groups of characteristics, as specified in the figure. In panel (a), each FN's effectiveness is correlated with the characteristics of her modal supervisor. In panel (b), team characteristics are constructed from FNs who overlap at the same site as the focal FN, weighted by the number of months of overlap. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Table A.1: Summary statistics of mother characteristics in different analytical samples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All mothers		All mothers with valid birth outcome			All mothers with age 2 outcomes		
	Mean	Std. Dev	Mean	Std. Dev	p-val (1) = (3)	Mean	Std. Dev	p-val (1) = (6)
Demographics								
Age (years)	214.6	(16.324)	214.4	(16.304)	0.211	214.0	(15.951)	0.001
Gestational age (weeks)	17.9	(4.905)	18.0	(4.879)	0.239	17.8	(4.798)	0.026
Primary language English	0.955	(0.208)	0.958	(0.201)	0.109	0.965	(0.184)	0.000
Ethnicity: Black	0.054	(0.226)	0.054	(0.226)	0.942	0.051	(0.221)	0.305
Ethnicity: Asian	0.026	(0.159)	0.025	(0.157)	0.580	0.024	(0.153)	0.233
Ethnicity: Mixed/Other	0.067	(0.250)	0.066	(0.248)	0.657	0.061	(0.239)	0.031
Social and family indicators								
Mother has partner	0.767	(0.423)	0.766	(0.423)	0.922	0.777	(0.417)	0.037
Contact with bio dad: Never	0.139	(0.346)	0.138	(0.345)	0.692	0.129	(0.336)	0.012
Contact with bio dad: < Weekly	0.058	(0.234)	0.058	(0.235)	0.861	0.054	(0.227)	0.172
Contact with bio dad: Weekly	0.108	(0.310)	0.108	(0.311)	0.885	0.104	(0.305)	0.250
Contact with bio dad: Daily	0.477	(0.499)	0.480	(0.500)	0.462	0.498	(0.500)	0.000
Lives with own partner	0.283	(0.450)	0.279	(0.449)	0.405	0.284	(0.451)	0.885
Lives with own mother	0.541	(0.498)	0.552	(0.497)	0.029	0.566	(0.496)	0.000
Economic indicators								
NEET	0.532	(0.499)	0.527	(0.499)	0.305	0.504	(0.500)	0.000
Lacks good qualifications	0.689	(0.463)	0.689	(0.463)	0.994	0.679	(0.467)	0.073
Any vocational qualification	0.427	(0.495)	0.427	(0.495)	0.986	0.440	(0.496)	0.026
Ever in paid work	0.444	(0.497)	0.442	(0.497)	0.685	0.456	(0.498)	0.049
Current in paid work	0.169	(0.374)	0.169	(0.375)	0.911	0.178	(0.382)	0.036
Income only from benefits	0.332	(0.471)	0.329	(0.470)	0.554	0.320	(0.467)	0.038
Residence: Private renter	0.212	(0.409)	0.209	(0.407)	0.382	0.207	(0.405)	0.241
Residence: Social renter	0.459	(0.498)	0.462	(0.499)	0.500	0.470	(0.499)	0.046
Residence: Non-traditional	0.074	(0.262)	0.070	(0.255)	0.150	0.054	(0.226)	0.000
Residence: Non-response	0.144	(0.351)	0.139	(0.346)	0.178	0.120	(0.325)	0.000
Social services								
Any social care	0.161	(0.367)	0.157	(0.364)	0.255	0.123	(0.329)	0.000
Any social services	0.504	(0.500)	0.506	(0.500)	0.665	0.498	(0.500)	0.269

Note: Summary statistics of characteristics of mothers enrolling in the FNP between 2007 and 2013. Columns 1 and 2 refer to the full sample of mothers enrolling in this period. Columns 3 and 4 restrict to our birth outcomes analytical sample (including conditioning on birth outcomes being observed and on being assigned to a family nurse with at least four other valid clients). Columns 6 and 7 refer to families in our analytical sample with at least one outcome observed at age 2. Sample size varies slightly across variables, between 22,070 and 23,702 for the full sample; between 18,639 and 19,370 in the birth outcomes sample; and between 10,723 and 11,119 for the age 2 sample. Missingness is highest for the 'income only from benefits' and 'mother lacks good qualifications' indicators; outside of these, minimum sample sizes are 22,789; 19,226; and 11,023, respectively.

Table A.2: Test of within-site quasi-random assignment on analytical samples

	# outcomes	# of outcomes where mother Xs significant		
		No MHT	HBS	RW
Panel A: Birth outcome sample				
All family nurse characteristics	38	6	0	0
% of all family nurse characteristics		15.8%	0.0%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	1	0	0
Family nurse training	5	1	0	0
Process quality	11	4	0	0
Panel B: Cognitive development age 1 sample				
All family nurse characteristics	38	4	0	0
% of all family nurse characteristics		10.5%	0.0%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	1	0	0
Family nurse training	5	1	0	0
Process quality	11	2	0	0
Panel C: Socio-emotional development age 1 sample				
All family nurse characteristics	38	6	0	0
% of all family nurse characteristics		15.8%	0.0%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	1	0	0
Family nurse experience	5	1	0	0
Family nurse training	5	1	0	0
Process quality	11	3	0	0

Note: The table reports the results of the test of quasi-random assignment of clients to family nurses on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the family nurse characteristics in a regression of a family nurse characteristic on mother characteristics not known at the time of assignment, controlling for mother characteristics known at the time of assignment (age, gestational age, language, and ethnicity), the family nurse's caseload size and complexity at the time of assignment, site fixed effects, and mother-year enrolment dummies. Standard errors are clustered at the site level in these regressions. The table reports the number of family nurse characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test where the p-value is not adjusted for multiple hypothesis testing in column (2), where it is adjusted using the Holm-Bonferroni adjustment in column (3), and where it is adjusted using the Romano and Wolf procedure in column (4). Column (1) reports the total number of family nurse characteristics. In each panel, the first two rows refer to the entire set of 38 family nurse characteristics we test for, while the rows below report the results by groups of family nurse characteristics.

Table A.3: Test of within-site quasi-random assignment on analytical samples (cont.)

	# outcomes # outcomes	# of outcomes where mother Xs significant		
		No MHT	HBS	RW
Panel D: Cognitive development age 2 sample				
All family nurse characteristics	38	2	0	0
% of all family nurse characteristics		5.3%	0.0%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	1	0	0
Family nurse training	5	0	0	0
Process quality	11	1	0	0
Panel E: Socio-emotional development age 2 sample				
All family nurse characteristics	38	2	1	0
% of all family nurse characteristics		5.3%	2.6%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	1	1	0
Family nurse training	5	0	0	0
Process quality	11	1	0	0

Note: The table reports the results of the test of quasi-random assignment of clients to nurses on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the nurse characteristics in a regression of a family nurse characteristic on mother characteristics not known at the time of assignment, controlling for mother characteristics known at the time of assignment (age, gestational age, language, and ethnicity), the family nurse's caseload size and complexity at the time of assignment, site fixed effects, and mother-year enrolment dummies. Standard errors are clustered at the site level in these regressions. The table reports the number of family nurse characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test where the p-value is not adjusted for multiple hypothesis testing in column (2), where it is adjusted using the Holm-Bonferroni adjustment in column (3), and where it is adjusted using the Romano and Wolf procedure in column (4). Column (1) reports the total number of family nurse characteristics. In each panel, the first two rows refer to the entire set of 38 family nurse characteristics we test this for, while the rows below report the results by groups of family nurse characteristics.

Table A.4: Test of across-site quasi-random assignment on sample of mothers enrolled in the FNP between 2007 and 2013

	# outcomes	# of outcomes where mother Xs significant		
		No MHT	HBS	RW
A - Benchmark specification				
All family nurse characteristics	38	4	2	0
% of all family nurse characteristics		10.5%	5.4%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	4	2	0
B - Not controlling for case complexity				
All family nurse characteristics	38	6	2	0
% of all family nurse characteristics		15.8%	5.3%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	6	2	0
C - Not controlling for case complexity or mother basic referral variables				
All family nurse characteristics	38	6	4	1
% of all family nurse characteristics		15.8%	10.5%	2.6%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	6	4	1

Note: The table reports the results of the test of quasi-random assignment of clients to family nurses on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the family nurse characteristics in a regression of a family nurse characteristic on mother characteristics not known at the time of assignment, controlling for mother characteristics known at the time of assignment (age, gestational age, language, and ethnicity), the family nurse's caseload size and complexity at the time of assignment, and mother-year enrolment dummies. Standard errors are clustered at the site level in these regressions. The table reports the number of family nurse characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test where the p-value is not adjusted for multiple hypothesis testing in column (2), where it is adjusted using the Holm-Bonferroni adjustment in column (3), and where it is adjusted using the Romano and Wolf procedure in column (4). Column (1) reports the total number of family nurse characteristics. In each panel, the first two rows refer to the entire set of 38 family nurse characteristics we test this for, while the rows below report the results by groups of family nurse characteristics.

Table A.5: Test of across-site quasi-random assignment on analytical samples

	# outcomes	# of outcomes where mother Xs significant		
		No MHT	HBS	RW
Panel A: Birth outcome sample				
All family nurse characteristics	38	4	2	0
% of all family nurse characteristics		10.5%	5.3%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	4	2	0
Panel B: Cognitive development age 1 sample				
All family nurse characteristics	38	4	1	0
% of all family nurse characteristics		10.5%	2.6%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	4	1	0
Panel C: Socio-emotional development age 1 sample				
All family nurse characteristics	38	5	1	0
% of all family nurse characteristics		13.2%	2.6%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	5	1	0

Note: The table reports the results of the test of quasi-random assignment of clients to nurses on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the family nurse characteristics in a regression of a family nurse characteristic on mother characteristics not known at the time of assignment, controlling for mother characteristics known at the time of assignment (age, gestational age, language, and ethnicity), the family nurse's caseload size and complexity at the time of assignment, and mother-year enrolment dummies. Standard errors are clustered at the site level in these regressions. The table reports the number of family nurse characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test where the p-value is not adjusted for multiple hypothesis testing in column (2), where it is adjusted using the Holm-Bonferroni adjustment in column (3), and where it is adjusted using the Romano and Wolf procedure in column (4). Column (1) reports the total number of family nurse characteristics. In each panel, the first two rows refer to the entire set of 38 family nurse characteristics we test this for, while the rows below report the results by groups of family nurse characteristics.

Table A.6: Test of across-site quasi-random assignment on analytical samples (cont.)

	# outcomes	# of outcomes where mother Xs significant		
		No MHT	HBS	RW
Panel D: Cognitive development age 2 sample				
All family nurse characteristics	38	3	1	0
% of all family nurse characteristics		7.9%	2.6%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	3	1	0
Panel E: Socio-emotional development age 2 sample				
All family nurse characteristics	38	3	1	0
% of all family nurse characteristics		7.9%	2.6%	0.0%
Family nurse demographics	5	0	0	0
Family nurse qualifications	12	0	0	0
Family nurse experience	5	0	0	0
Family nurse training	5	0	0	0
Process quality	11	3	1	0

Note: The table reports the results of the test of quasi-random assignment of clients to nurses on the full sample of clients enrolled in the FNP between 2007 and 2013. We test whether the coefficients on mother characteristics not known at the time of assignment are jointly significant in predicting the family nurse characteristics in a regression of a family nurse characteristic on mother characteristics not known at the time of assignment, controlling for mother characteristics known at the time of assignment (age, gestational age, language, and ethnicity), the family nurse's caseload size and complexity at the time of assignment, and mother-year enrolment dummies. Standard errors are clustered at the site level in these regressions. The table reports the number of family nurse characteristics for which the null that the coefficients on client characteristics are jointly 0 using a F-test where the p-value is not adjusted for multiple hypothesis testing in column (2), where it is adjusted using the Holm-Bonferroni adjustment in column (3), and where it is adjusted using the Romano and Wolf procedure in column (4). Column (1) reports the total number of family nurse characteristics. In each panel, the first two rows refer to the entire set of 38 family nurse characteristics we test this for, while the rows below report the results by groups of family nurse characteristics.

Table A.7: Validation of leave-out estimates of family nurse effectiveness

	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
Panel A: Birth outcomes									
Leave-out nurse FE	0.123	0.157	0.179	0.186					
	(0.066)	(0.064)	(0.065)	(0.064)					
N	18,208	18,208	18,208	18,208					
R^2	0.000	0.071	0.077	0.084					
p-val: $H_0 = 1$	0.000	0.000	0.000	0.000					
Panel B: Age 1 cognitive					**Panel C: Age 2 cognitive**				
Leave-out nurse FE	0.834	0.838	0.848	0.867	Leave-out nurse FE	0.728	0.739	0.744	0.757
	(0.039)	(0.039)	(0.041)	(0.039)		(0.049)	(0.049)	(0.051)	(0.048)
N	13,755	13,755	13,755	13,755	N	10,291	10,291	10,291	10,291
R^2	0.034	0.041	0.047	0.078	R^2	0.022	0.031	0.037	0.066
p-val: $H_0 = 1$	0.000	0.000	0.000	0.001	p-val: $H_0 = 1$	0.000	0.000	0.000	0.000
Panel D: Age 1 socio-emotional					**Panel E: Age 2 socio-emotional**				
Leave-out nurse FE	0.891	0.895	0.902	0.909	Leave-out nurse FE	0.747	0.738	0.750	0.768
	(0.032)	(0.033)	(0.031)	(0.031)		(0.044)	(0.044)	(0.045)	(0.046)
N	12,988	12,988	12,988	12,988	N	9,596	9,596	9,596	9,596
R^2	0.075	0.078	0.093	0.138	R^2	0.031	0.034	0.046	0.086
p-val: $H_0 = 1$	0.001	0.002	0.001	0.003	p-val: $H_0 = 1$	0.000	0.000	0.000	0.000
Basic controls		Y	Y	Y			Y	Y	Y
Time FE			Y	Y				Y	Y
Site FE				Y					Y

Note: The table reports the coefficient on $\hat{\mu}_{n,-i}$, the family nurse effectiveness estimated via a 'leave-out' method. We estimate these leave-out measures on outcomes demeaned within site, as in our main specification). We then regress outcomes on these leave-out measures of FN effectiveness. Source: Authors' calculations using data from the Department of Health and Social Care.

B Measures of caseload complexity

As we outline in section 2, FNP supervisors consider their team's relative caseloads when allocating new mothers to a family nurse. In our discussions with FNP administrators, they stressed that their assessment of caseload was not purely a numeric one: they also considered caseload 'complexity' and support needs. Of course, at the time of assignment, the new mother's own complexity was not known to the FNP programme: only basic referral details were available. However, supervisors were aware of the relative complexity of their team's *existing* caseloads, and used this to assess which family nurses had more capacity to take on a new client.

In our main specification (Equation 3), we control for two measures of the assigned family nurse's caseload, relative to other family nurses in her site at that time. The first is a simple count of her number of clients, i.e. the *size* of her caseload: we calculate the total number of clients of each family nurse in each month, and standardise within site-month to get a measure of relative caseload size. The second is a measure of caseload *complexity*, that we construct in two steps. In a first step, we create a complexity index for each mother in the FNP, based on the extended characteristics reported during her enrolment interview. Specifically, we include:

- Basic demographics: Age in months at enrollment; ethnicity; whether speaks English.
- Social and family indicators: Partnership status; frequency of contact with child's biological father; living with own partner; living with own mother.
- Economic indicators: Not in education, employment or training (NEET); lacks good qualifications; currently in education; any vocational qualification; ever in paid work; currently in paid work; income only from benefits; housing tenure (owner-occupied, private renter, social renter, non-traditional accommodation such as jail or foster care, non-response).
- Social services: Any social services; any social care.
- Infant feeding plans: Definitely planning not to breastfeed; maybe planning to breastfeed.
- Obstetric history: Gestational age at enrolment; gestational age at the start of antenatal care; number of previous pregnancies.
- Health history: Number of pre-existing health conditions; number of urinary tract infections since discovering pregnancy; number of treatments for sexually transmitted infections since discovering pregnancy; Body Mass Index.
- Risky behaviours: Any smoking during pregnancy; number of cigarettes smoked in the last 48 hours; units of alcohol consumed in the last two weeks; number of marijuana pipes/spliffs smoked in the last two weeks; number of times using street drugs (including cocaine and ecstasy) in the last two weeks.
- Mental health: Score on a mental health screening tool; score on a locus of control screening

tool.

- History of abuse: Ever experienced abuse (from a partner or other important person); number of times physically abused since becoming pregnant; number of times subjected to forced sexual relations in the last year; number of instances that abuse resulted in burns, severe bruises or broken bones in the last year; number of times abuse resulted in head, internal, or permanent injuries in the last year; whether currently afraid of partner/ex-partner/any important person.
- Recruitment: Number of visits required to recruit into the FNP.

We aggregate these indicators using factor analysis, accounting for both continuous and discrete measures.

In a second step, we construct a monthly panel of each family nurse's caseload. We do so by first taking a simple average of the case complexity factors across the family nurse's 'active' clients that month (i.e. those who are still receiving visits), and then by standardising this average within site and month. In this way, we obtain a measure of the focal family nurse's caseload complexity relative to her colleagues.*

Finally, since our measures of relative caseload size and complexity are standardised within site-month, we also control for the number of family nurses active within the site-month. These three variables - relative caseload size, relative caseload complexity, and number of active FNs in the site-month - make up the vector of 'caseload complexity' controls ($C_{nst_{0i}}$) in Equation 3.

*In cases where any indicator is missing, we impute it with a constant and include a binary variable for missing indicator.

C Simulations to validate approach to multiple hypothesis testing

In subsection 4.3, we provide evidence supporting the quasi-random assignment of clients to family nurses. Here we detail our adaptation of two standard MHT algorithms — the Holm-Bonferroni-Sidak (HBS) (Holm, 1979) adjustment, and the Romano-Wolf (RW) (Romano and Wolf, 2005) stepdown procedure — to the case where we have multiple predictors of interest as well as multiple outcomes. Both algorithms are conventionally based on t-statistics; we adapt them to use instead the F-statistic from a joint test of the 19 mother characteristics in each regression, which tests whether mothers' characteristics jointly predict each family nurse characteristic.

C.1 Description of simulated datasets

To assess the performance of our adapted algorithm, we run 500 permutations of each of four simulation scenarios designed to span a range of known dependency levels between mother and family nurse characteristics:

- Simulation 1: Random assignment of mothers to family nurses.
- Simulation 2: Synthetic mothers with randomly generated characteristics.
- Simulation 3: Synthetic mothers with population-level dependency on family nurse characteristics.
- Simulation 4: Synthetic mothers with high dependency on family nurse characteristics.

In Simulation 1, we randomly assign mothers to family nurses (i.e., entirely unconditionally). This simulation preserves the marginal distribution of both family nurse and mother characteristics in our sample, but breaks any systematic relationship between them. Each of the 500 permutations draws a new random assignment, providing a benchmark distribution of test statistics under the null of no relationship.

In Simulations 2-4, we generate synthetic mother characteristics as a function of family nurse characteristics and a random shock, creating as many synthetic mothers per family nurse as she has actual clients. Specifically, for each mother characteristic C_i^m, $m = 1, \ldots, 19$, we estimate the following regression on the characteristics Z_n of the family nurse assigned to mother i:[†]

$$C_i^m = \alpha^m + \beta^m Z_n + \nu_i^m \tag{7}$$

yielding a vector of estimated coefficients $\hat{\beta}^m$ — capturing the relationship between each mother characteristic and family nurse characteristics in the data — and a vector of residuals $\hat{\nu}_i^m$. We then

[†]Note that, for simplicity, we omit the site and time fixed effects as well as the case complexity variables included in our main random assignment testing in Equation 5.

generate synthetic mother characteristics as:

$$\widetilde{C}_i^m = \Omega \cdot \hat{\beta}^m Z_n + \hat{\epsilon}_i^m \tag{8}$$

where $\hat{\epsilon}_i^m$ is drawn from a multivariate normal distribution with mean zero and variance-covariance matrix equal to that of the estimated residuals $\hat{\nu}_i^m$. The scalar $\Omega \in \{0, 1, 5\}$ governs the degree of built-in dependence between synthetic mother characteristics and family nurse characteristics:

- Simulation 2: $\Omega = 0$ (synthetic "mothers" are based only on random shocks).
- Simulation 3: $\Omega = 1$ (synthetic "mother" traits exhibit a degree of dependence on family nurse characteristics matching that observed in the data).
- Simulation 4: $\Omega = 5$ (synthetic "mother" traits include a high level of dependency on family nurse characteristics).

For each simulation, we run 500 permutations. In each permutation $p \in \{1, \ldots, 500\}$ of Simulations 2–4, we draw a new shock vector $\hat{\epsilon}_{i,p}^m$ from a multivariate normal distribution with mean zero and variance-covariance matrix equal to that of the estimated residuals $\hat{\nu}_i^m$.

C.2 Simulation results

In Figure C.19 and Figure C.20, we summarise the results of this simulation exercise. Each column represents one of the four simulations, as described above. We run 500 permutations of each simulation, in each case running 38 regressions (one for each of the family nurse characteristics considered). For each permutation, we calculate the share of the 38 regressions where the synthetic "mother" characteristics are jointly significant, after applying the relevant MHT algorithm.

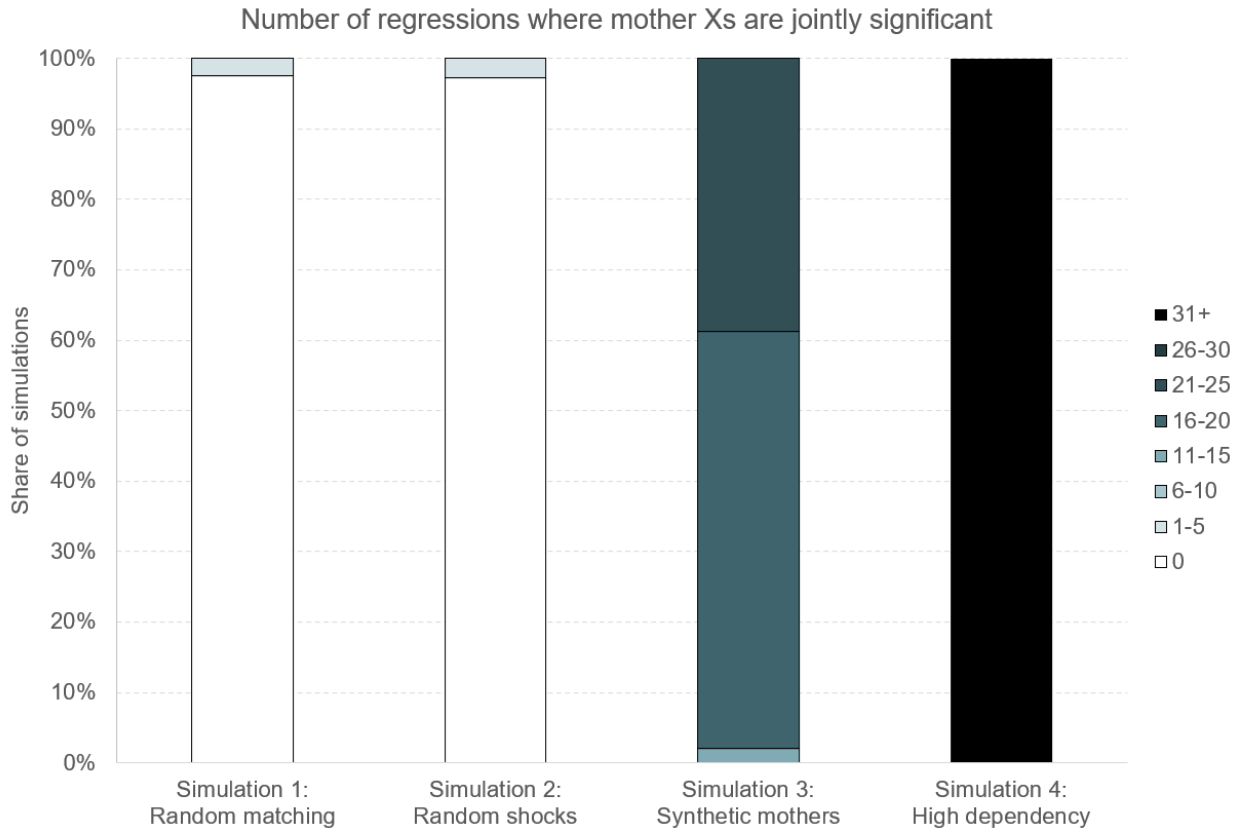
As expected, both algorithms detect very few significant effects in Simulations 1 and 2, where "mothers" are either randomly assigned to family nurses or have characteristics that are randomly generated. For example, of the 500 permutations of Simulation 1 run using the Holm-Bonferroni algorithm, just 12 had any jointly significant mother characteristics. In each of these 12 cases, only one of the 38 family nurse characteristics regressions had jointly significant mother predictors. This is reassuring evidence of a low false positive rate.

As we increase the level of dependence between the family nurse characteristics and the characteristics of the simulated "mothers", both algorithms are much more likely to detect significant relationships between mother characteristics and family nurse traits. For example, all of the 500 permutations in Simulation 4 using the Holm-Bonferroni algorithm found that "mother" characteristics significantly predicted more than 30 out of the 38 family nurse characteristics. This provides evidence that these algorithms do detect higher levels of dependence.

Notably, we find that the Holm-Bonferroni algorithm is more sensitive than the Romano-Wolf algorithm: it is more likely to detect significant relationships in Simulations 3 and 4. This is in

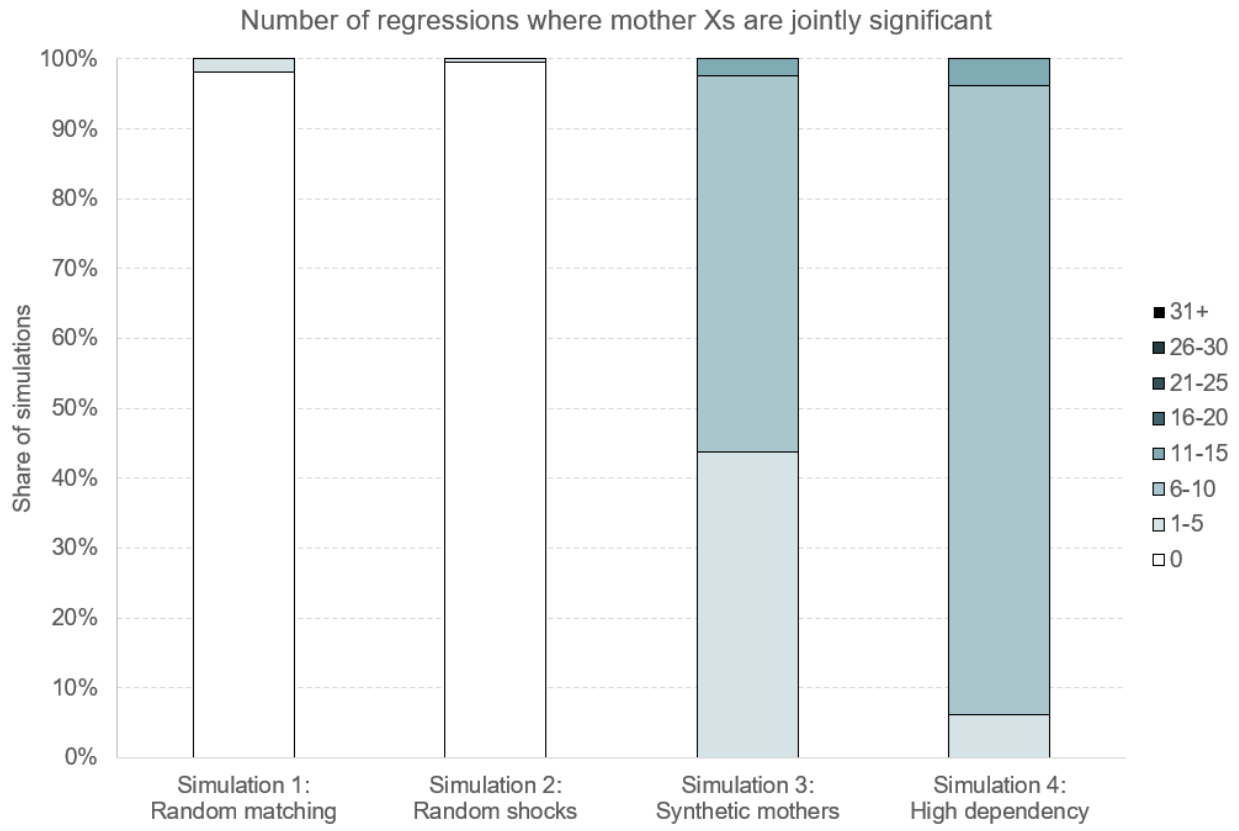
line with the existing literature, which shows that the Romano-Wolf algorithm is more conservative (i.e. less likely to reject a null hypothesis of no relationship) (Romano and Wolf, 2005). In our application, the Holm-Bonferroni algorithm therefore constitutes the more demanding test of quasi-random assignment.

Figure C.19: Distribution of significance in simulated datasets, applying the modified Holm-Bonferroni algorithm



Note: The figure shows the number of regressions in which "mother" and "family nurse" characteristics were statistically significantly related, under four simulations with progressively greater levels of in-built dependence. Source: Authors' calculations using data provided by the Department for Health and Social Care.

Figure C.20: Distribution of significance in simulated datasets, applying the modified Romano-Wolf algorithm



Note: The figure shows the number of regressions in which "mother" and "family nurse" characteristics were statistically significantly related, under four simulations with progressively greater levels of in-built dependence. Source: Authors' calculations using data provided by the Department for Health and Social Care.

D Family nurse effectiveness at retaining mothers in the FNP programme

subsection 4.4 documents that mothers can choose to disengage from the FNP programme: a third of them had left by the child's first birthday, rising to half by the end of the intervention at the child's second birthday. While there is little evidence of economically meaningful differences in attrition by mothers' socio-economic characteristics, a concern remains that some family nurses may be more effective at retaining clients, leading to a more negatively-selected caseload – and hence attenuating their estimated effectiveness at improving child development. In subsection 5.2, we address this concern directly by estimating family nurse effectiveness on a common sample of mothers who are retained until the end of the programme, finding results very similar to our main estimates.

This appendix presents a complementary analysis focused on retention. We define the outcome as the number of months a mother remains in the programme, standardised to have mean zero and standard deviation one, and estimate Equation 3 to recover FN effectiveness at promoting retention.

Table D.8 presents the summary statistics for the variation in family nurse effectiveness at retaining mothers, alongside the main results on child outcomes presented in Table 4. Figure D.21 plots the distribution of family nurse fixed effects for retention and for our five main outcomes. Overall, we find a similar degree of variation in effectiveness for retention as for other outcomes (a one-standard deviation increase in family nurse effectiveness leads to a 0.20SD increase in retention, or about two additional months in the programme).

Given variation in family nurse effectiveness at retaining mothers, a key question is whether family nurses who are better at retaining mothers also appear to be systematically worse (or better) on our measures of effectiveness for child development outcomes. Table D.9 shows there are no significant correlations between family nurse effectiveness at retaining clients on the one hand, and effectiveness at promoting children's outcomes on the other.

A final question of interest is what characteristics predict family nurses who are particularly effective at retaining mothers. Figure D.22 shows that the demographic characteristics of the family nurse are predictive of retention: supervisors and non-white nurses tend to see clients leave the programme earlier. Also, family nurse-mother matched characteristics are predictive of retention: more specifically, mothers whose ethnicity matches that of their family nurse are more likely to remain in the FNP for longer (although younger teens are more likely to attrit when paired with a family nurse with specialist experience in working with teenagers). This is in sharp contrast to the results presented in Figure 3, which showed little role for demographics in predicting FN effectiveness at improving cognitive development and a negative correlation to ethnic match.

Table D.8: Distribution of family nurse effectiveness for different child outcomes

		Child outcome				
	Retention	Birth outcomes	Cognitive age 1	Cognitive age 2	Socio-emo. age 1	Socio-emo. age 2
Panel A - Within sites						
SD	0.200	0.109	0.218	0.196	0.307	0.234
p90 - p10	0.518	0.288	0.547	0.488	0.761	0.588
p75 - p25	0.274	0.141	0.279	0.232	0.37	0.307
Panel B - Across sites						
SD	0.243	0.121	0.266	0.239	0.371	0.292
p90 - p10	0.634	0.311	0.682	0.628	0.908	0.75
p75 - p25	0.321	0.154	0.363	0.298	0.46	0.4
No. of FNs	660	647	616	585	611	574
No. of mothers	21,236	18,258	13,820	10,383	13,060	9,700

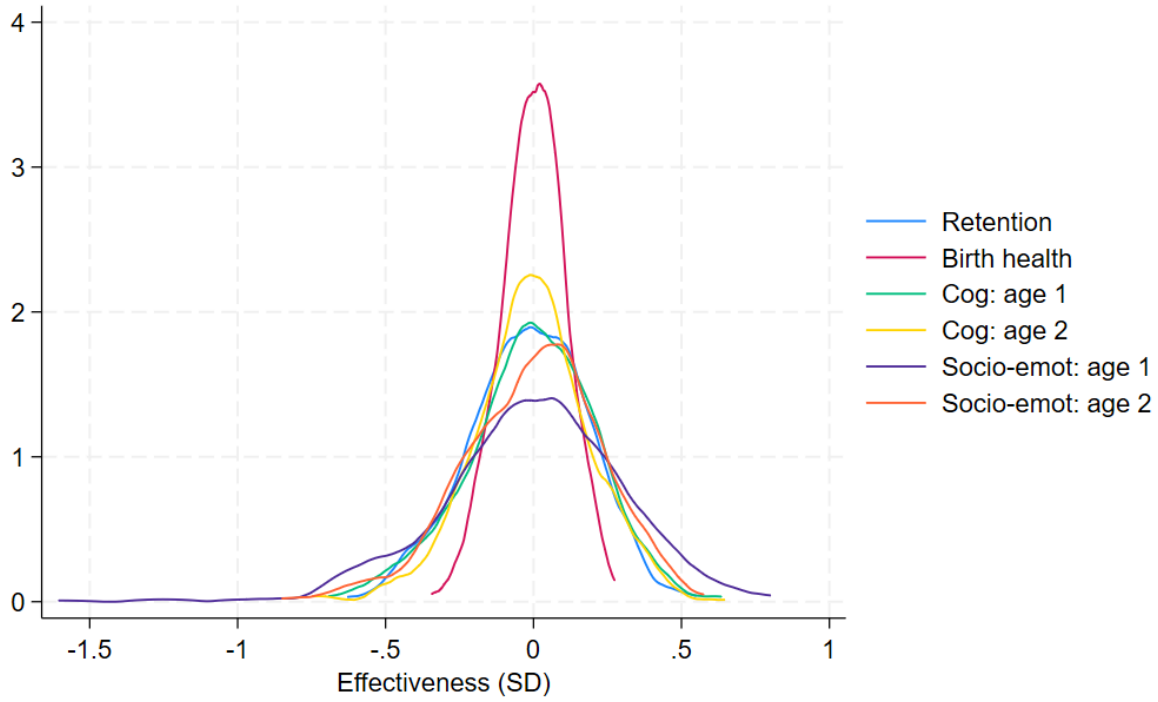
Note: The table reports the standard deviation of the distribution of family nurse fixed effects, estimated as in Equation 3. The sample includes mothers who joined the FNP by December 2013 and have a valid outcome measure, and FNs who have at least five mothers. In Panel A, outcomes are first demeaned within site to account for fixed differences across FNP sites. We apply a Bayesian shrinkage estimator to all results. Source: Authors' calculations using data from the Department of Health and Social Care.

Table D.9: Correlations between family nurse effectiveness at promoting different child outcomes

	Retention (months) (1)	Birth (2)	Cognitive age 1 (3)	Cognitive age 2 (4)	Socio-emotional age 1 (5)	Socio-emotional age 2 (6)
(1) Retention (mths.)		0.073	0.041	-0.009	0.026	0.049
(2) Birth outcomes	0.073		**0.165**	**0.143**	0.004	0.067
(3) Cognitive, age 1	0.041	**0.165**		**0.671**	**0.338**	**0.338**
(4) Cognitive, age 2	-0.009	**0.143**	**0.671**		**0.254**	**0.373**
(5) Socio-emot., age 1	0.026	0.004	**0.338**	**0.254**		**0.575**
(6) Socio-emot., age 2	0.049	0.067	**0.338**	**0.373**	**0.575**	

Note: Correlations in **bold** are statistically significant at the 5% level.

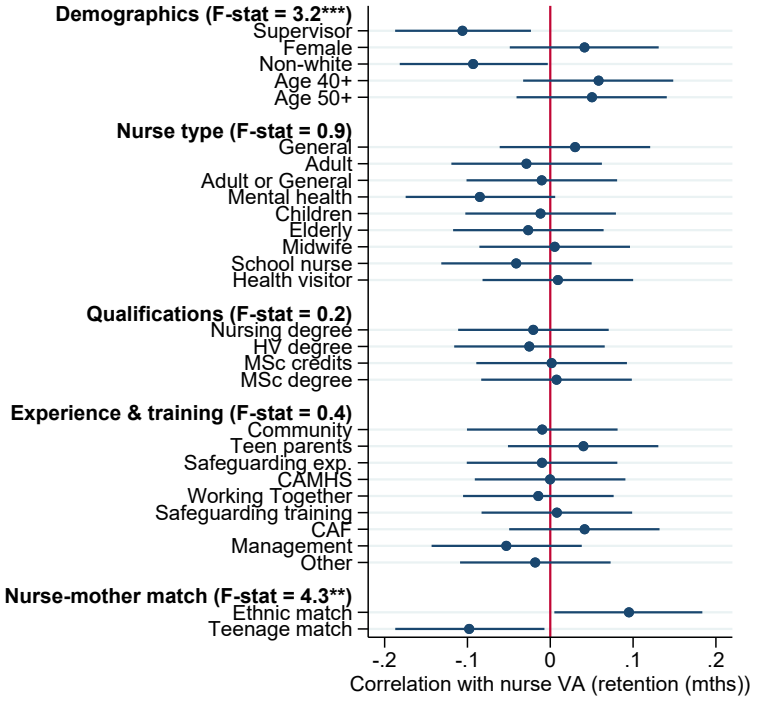
Figure D.21: Distribution of family nurse effectiveness



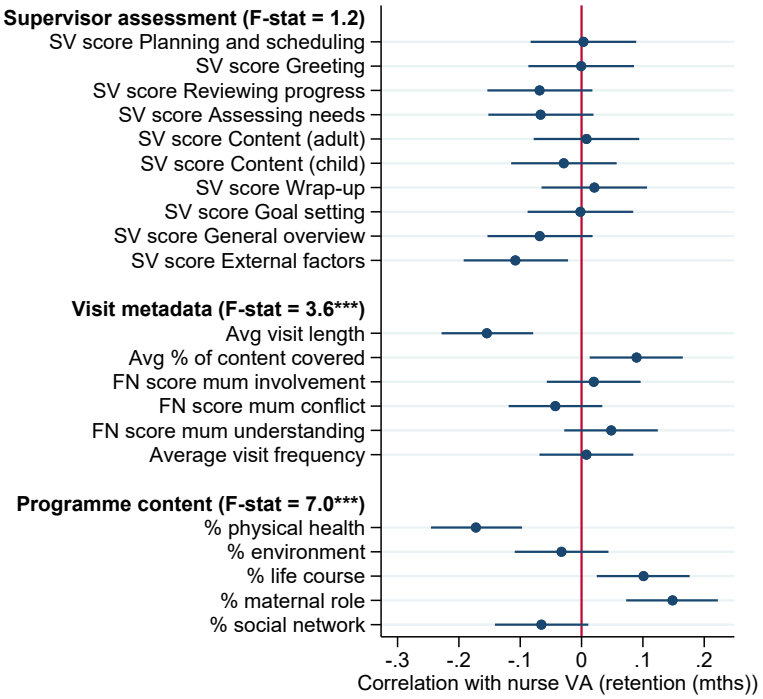
Note: See Table 4.

Figure D.22: Pairwise correlations between family nurse effectiveness at promoting client retention and FN quality indicators

(a) FN 'hiring' characteristics (structural quality indicators)



(b) FN process quality measures



Note: The figure shows the point estimate and 95% confidence intervals for the pairwise correlations between family nurse effectiveness and each FN quality indicator, as specified. The F-statistic is based on a regression of family nurse effectiveness on groups of characteristics, as specified in the figure. Due to missingness in FN characteristics, these results are based on 419 nurses. Source: Authors' calculations using data provided by the Department for Health and Social Care.

E Constructing measures of potential mechanisms

The Family Nurse Partnership aims to provide a holistic intervention to support mothers and children. Programme content covers mothers' physical health and risky behaviours; their 'life course', including education and employment; their social network, including safe and unsafe relationships; their understanding of, and ability to, support child development; and the safety and healthiness of their child's environment.

Given the wide range of programme content, the FNP Information System contains a wealth of information about specific indicators for risky behaviours, socio-economic circumstances, maternal relationships, maternal mental health, and investments in children. We view these as potential mechanisms through which effective family nurses support children's cognitive and socio-emotional development.

We use factor analysis to aggregate different specific indicators into five broader measures, loosely mapping into the different dimensions of the FNP curriculum. Where data allows, we construct two measures for each domain: one up to and including the child's age 1, the other taking in all information until the child's age 2. Since there is significant path dependence in each of these mechanisms, we include the full history of available information in each of our factors to capture the potential for effective family nurses to intervene and shift behaviour at different points of the life cycle. We summarise the indicators used to construct each factor in Figure E.23. Full results from our factor analysis are available on request.

Figure E.23: Indicators included in each potential mechanism

Mechanism	36 weeks' pregnancy	Birth	Child age 1	Child age 2
Risky behaviours	# cigarettes in last 2 days		# cigarettes in last 2 days	
	Any alcohol in last 2 weeks		Any alcohol in last 2 weeks	
	Any marijuana in last 2 weeks		Any marijuana in last 2 weeks	
		# UTI treatments since intake		
		Any STI treatment since intake		
			Any unprotected sex in last 6 months	Any unprotected sex in last 6 months
Socio-economic circumstances			Not in employment/educ./training	Not in employment/educ./training
			Income solely from welfare	Income solely from welfare
			# of GCSEs attained	# of GCSEs attained
Relationships		Any abuse	Any abuse	
			Partnered	Partnered
			Living with own mother	Living with own mother
			Frequency biological dad cares for child	Frequency biological dad cares for child
Mental health	HADS: Anxiety			
	HADS: Depression			
				Positive affect
				Locus of control
Investments in child			All vaccinations up-to-date	
			Weeks exclusively breastfed	
			Use of Sure Start services	

Note: Our 'age 1' mechanisms contain all information to the left of the vertical line; our 'age 2' mechanisms contain the full set of information.